AFRL-SR-BL-TR-98

REPORT DOCUMENTATION PAGE

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	30 Nov 96	FINAL TECH RPT, 01 OCT 95 TO 31 MAY 96

4. TITLE AND SUBTITLE
An Active Vision Based SAR-FLIR Fusion ATR System for Detection and Recognition of Ground Targets

5. FUNDING NUMBERS
F49620-95-C-0079

6. AUTHOR(S)
Drs. S. Raghavan, R. Poovendran, S. Srinivasan and L. Kanal

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
LNK Corporation, Inc.
Riverdale MD 20737

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR/NM
110 Duncan Avenue Suite B115
Bolling AFB DC 20332-8050

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT
Unlimited Distribution

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

1) A constant False Alarm Rate (CFAR) has been developed and implemented for real-time performance on an inexpensive Single Instruction Multiple Data (SIMD) hardware. 2) An innovative registration algorithm has been developed to register the SAR-passive image pair. 3) A model-based recognition scheme using a model-image alignment approach has been implemented. 4) An end-to-end demonstration of the proof-of-concept demonstration of the SAR-passive image fusion approach has been accomplished.

19980129 065

14. SUBJECT TERMS	15. NUMBER OF PAGES
detection, multi-sensor registration, recognition	61
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

An Active Vision Based SAR-FLIR Fusion ATR System for Detection and Recognition of Ground Targets

Item 002 AA
Final Technical Report

S. Raghavan, R. Poovendran, S. Srinivasan, and L. Kanal
LNK Corporation Inc., Riverdale, MD.

R. Chellappa and C. Shekhar
University of Maryland, College Park, MD.

STTR Phase I Contract
F49620-95-C-0079

Supported by
Air Force of Scientific Research

COTR: Dr. Abe Waksman

Principal Investigator:
Dr. Srinivasan Raghavan

Prime Contractor: LNK Corporation Inc.
6811, Kenilworth Avenue, Suite 306
Riverdale, MD 20737.
Tel: (301) 927 3223

Co-PI: Dr. Rama Chellappa
Sub-contractor: University of Maryland
College Park, MD.

UNCLASSIFIED

STTR RIGHTS LEGEND

These STTR data are furnished with STTR rights under contract no. **F49620-95-C-0079**. For a period of four (4) years after the acceptance of all items to be delivered under this contract, the Government agrees to use these data for Government purposes only, and they shall not be disclosed outside the Government (including disclosure for procurement purposes) during such period without permission of the contractor, except that, subject to the foregoing use and disclosure prohibitions, such data may be disclosed for use by support contractors. After the aforesaid four (4) year period the Government has a royalty free license to use, and to authorize use of this data by third parties. This Notice shall be affixed to any reproductions of these data, in whole or part.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Abe Waksman for his comments and suggestions throughout the period of performance of this contract. The authors also express their appreciation to all the LNK staff who contributed to this effort directly or indirectly.

TABLE OF CONTENTS

1. INTRODUCTION

The geo-political changes around the world in the last five years have clearly shown the need for the U.S. to maintain a strong technical edge in high-tech weaponry over potential adversaries. Unlike the thrust on ballistic missiles technology seen earlier, before the dissolution of Soviet Union, the recent thrust is more on accurate targeting. Targeting, particularly in the case of relocatable targets, is expected to pose a considerable challenge to future missions whether manned or unmanned. Relocatable targets such as mobile launchers, armored vehicles, ammunition and equipment carriers, supply trucks, and mobile radar units played a crucial role in the Gulf war by providing a difficult-to-eliminate capability of the adversary. Use of smart weapons in that war provided a substantial advantage to the Allied Forces. Although, this experience demonstrated the vast potential of high-tech weaponry and smart sensors, it also helped point out the deficiencies in current systems, in particular, the need for robust and reliable Automatic Target Recognition (ATR) technology to improve the strike accuracy.

The goal of robust ATR can be achieved if we can make use of multiple sensors in a complementary fashion. While sensors such as Synthetic Aperture Radar (SAR) which provide a wide area of coverage obviously suffer from poor resolution, sensors such as Forward Looking Infrared (FLIR) which provide a good resolution suffer from a narrow field of view. In addition to this, to facilitate the all-night-all-weather targeting capability, we need to make certain that optimal use is made of the strengths of different sensors. In this sense, use of SAR for detection of targets has become very popular. Since SAR suffers from poor resolution and lack of look-angle-independent features, we need to complement the SAR with FLIR or electro-optical (EO) sensors for accurate target recognition.

For over a decade, a number of ATR initiatives such as the Reconnaissance, Surveillance, and Target Acquisition (RSTA), and Moving and Stationary Target Acquisition and Recognition

(MSTAR) program funded by the Defense Advance Research Projects Agency (DARPA), have resulted in very mature image understanding and sensor fusion algorithms at academic institutions such as the University of Maryland, College Park, MD. This STTR Phase I is essentially a technology-transfer initiative to incorporate mature detection and multi-sensor registration techniques developed at the University of Maryland into a robust model-based ATR module developed at LNK.

This report is organized into eight sections. In section 2, we discuss the Phase I technical objectives and summarize our work in reaching these objectives. In Section 3, a conceptual framework for a SAR-FLIR multi-sensor fusion prototype system is discussed. A computationally elegant registration scheme is presented in Section 4. Section 5 presents a discussion on the Constant False Alarm Rate detection algorithms used in the Phase I effort. The model-based vision technique is described in Section 6. We discuss the results in Section 7 and lay the foundation for the Phase II work in Section 8.

2. A SUMMARY OF THE PHASE I WORK

Synthetic Aperture Radar (SAR) platforms have been found very useful for target detection applications due to their all-weather, and all-day-and-night capabilities. However, unlike Electro-Optical (EO) sensors, SAR imaging systems are limited by their lack of preserving irradiance characteristics of the terrain objects. Serious difficulties arise when we use SAR for recognition of objects because it is difficult to extract good geometric features from SAR imagery. Passive sensors, on the other hand, remove this difficulty at the cost of a very narrow field of view. Therefore, it makes all the more sense to use SAR to locate potential areas where targets are likely to be present or simply Focus of Attention (FOA), and direct the passive sensors to obtain high resolution target-specific details from the areas of FOA.

Our primary goal in this Phase I was to assemble a suite of sensor fusion techniques for SAR-passive imagery fusion based on the on-going ATR work at LNK and the University of Maryland. In reaching this goal, our major technical objectives were to:

(1) understand the platform characteristics (physics and geometry of imaging) of the sensors involved;

(2) develop the techniques for extracting FOA from the SAR imagery

(3) develop registration algorithms for SAR-FLIR mapping;

(4) implement the model-alignment algorithm for target recognition;

(5) demonstrate the proof-of-concept on SAR-FLIR data available inhouse; and

(6) evaluate the ATR performance characteristics.

We have devoted a lot of attention to meeting each of these objectives in this Phase I. Our concept of using the SAR as the cueing sensor for locating targets and recognizing the targets is shown in Figure 2.1. We have achieved an effective technology-transfer of matured algorithms for detection, multi-sensor registration and recognition. However, due to the lack of enough data, the sixth objective of evaluating performance characteristics needs additional attention in Phase II when we acquire more data sets. In essence, the following are the major contributions resulting from this Phase I effort.

- A Constant False Alarm Rate (CFAR) has been developed and implemented for real-time performance on an inexpensive Single Instruction Multiple Data (SIMD) hardware.

- An innovative registration algorithm has been developed to register the SAR-passive image pair.

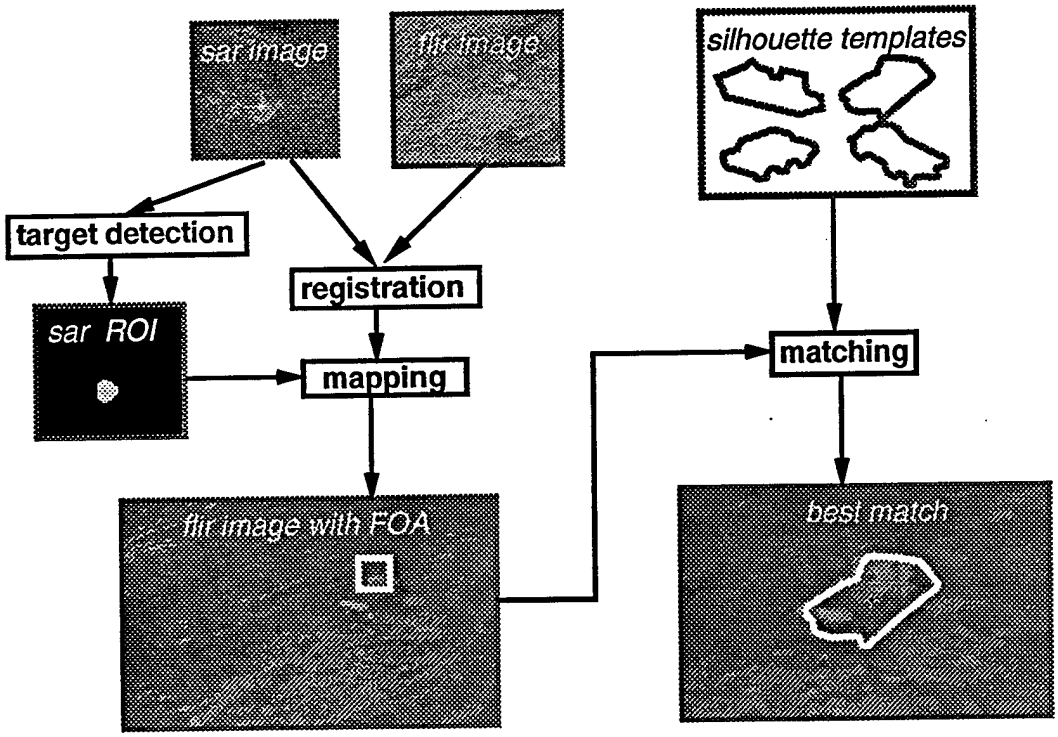

Figure 2.1: SAR based cueing of FLIR for detection and recognition of ground targets

- A model-based recognition scheme using a model-image alignment approach has been implemented.

- An end-to-end demonstration of the proof-of-concept demonstration of the SAR-passive image fusion approach has been accomplished.

We show the results of each of the three major steps, namely, detection, multi-sensor registration, and recognition, in Sections 4, 5, and 6. Our goal in Phase II is to develop a real-time prototype ATR system that exploits complementary strengths of SAR and passive imaging systems. The framework is a generic one in that we combine SAR with any type of passive sensor. In addition, if the resolution of the SAR is sufficiently high we will exploit the SAR data also for ATR in addition to using the SAR for detection and cueing. A more sophisticated SAR-passive image registration algorithm which takes into account imaging geometries of the active and passive sensors will also be the focus of the Phase II work. For improving the ATR performance of the Phase I approach, we will work on an indexing scheme for model selection based on geometric invariants.

3. BACKGROUND

Over the last few years, several approaches to sensor fusion have gained a strong momentum including Bayesian reasoning, Dempster-Shafer theory, fuzzy logic, blackboard frameworks, neural networks, and hybrid systems. Though our objective in writing this section is not to cover the entire spectrum of issues, we will attempt to provide a short summary of the earlier work in this section. A number of survey articles have been written in the recent past (e.g., Luo and Kay [1992]).

Multi-sensor fusion can be addressed at least at four different levels, namely, Level I, Level II, Level III, and Level IV. Consistent with the terminology adopted by the DARPA JDL sensor fusion group, the following connotation applies to these different levels of fusion. Level I fusion refers to fusion of information from similar sensor platforms. Level II fusion implies the integration of information from dissimilar sensors spatially and temporally separated. Level III fusion involves inference of individual activities of different entities in the theater. And Level IV fusion refers to the process of integrating and inferencing activities of multiple individual entities at the theater level. Our work in this context is a Level II fusion effort and therefore we will only discuss those efforts which involve Level II fusion.

One of the significant advantages of the multi-sensor based approach is the complementary information available in the different types of imagery to achieve segmentation of objects from the background. Camouflage that prevents segmentation in visual imagery may not be present in thermal imagery. For instance, Nandhakumar and Aggarwal [1988] presented a technique to segment thermal and visual images by matching corresponding features in the two images. Similarly, Tong et. al. [1988] developed a method to integrate Forward Looking Infra-Red (FLIR) sensors with Laser radar imagery to segment targets for target recognition.

Multi-sensor fusion has been often applied in the recent past to robotics in the context of navigation and recognition. Allen [1987] has developed a robotic system to recognize small objects using exploratory tactile sensing and three dimensional visual information. The system is model-based and components of the models are constructively added by combining tactile information and segmented surfaces of the objects. Earlier, Grimson and Lozano-Perez [1984] used model-based vision to fuse information from tactile and range sensors for measuring the position and computing the surface normals to identify and locate known objects. The Autonomous Land Vehicle (ALV) project at Carnegie Mellon University [Goto and Stenz 1987] is another example of using multiple

sensors including stereo cameras, laser range finders and sonar sensors for obstacle avoidance and outdoor navigation.

Neural networks also have played a critical role in sensor fusion research. For example, Rajapakse and Acharya [1990] proposed a hierarchical network to achieve fusion of images from multiple sensors for object classification. The hierarchical neural system based on the Neocognitron and ART 1, has many forward paths to carry information from different sensors and a fusion path to carry the combined information. The final decision is reached by looking at the information in the fusion path. They have not, however, reported experimentation with any real imagery. Also, the images from multiple sensors are assumed to have the same characteristics although with different scaling, rotation and noise. This may be an overly simplified representation of the real world.

Our work in sensor fusion began in 1990 with an SBIR effort for extraction of natural features from SAR and visual imagery [Raghavan et. al. 1991]. We used Gabor filters [Daugman 1988] to represent textures from co-registered multi-sensor imagery and neural networks to recognize the terrain features from normalized textural feature vectors. The neural networks are trained to classify the textural feature vectors from both radar and optical images into forests, water, and fields. We further developed man-made terrain feature extraction algorithms using a hybrid approach of traditional image understanding algorithms and fuzzy logic methods in another Navy SBIR Phase II effort [Raghavan et. al. 1996]. This Phase I STTR, in contrast with our earlier work, has been aimed at developing algorithms for SAR based cueing of passive imagery for target recognition.

4. IMAGE REGISTRATION

Registration of images is a task of fundamental importance in the field of image understanding (IU). The basic objective is to be able to relate the information in one image to the information in another. There are many applications where the two images are obtained using two different types of sensors. In remote-sensing and image exploitation (IE) applications, the sensors of interest are Synthetic Aperture Radar (SAR), Forward-Looking Infra-Red (FLIR) Electro-Optic (EO), and interferometric SAR (IFSAR). These sensors respond to scene characteristics in different, and often complementary, ways. It is therefore beneficial to use data from more than one type of sensor in any IU application. This type of sensor integration is possible only if the two images are co-aligned, or ``registered" with respect to a common coordinate system. Of course, for this to be meaningful, the images should share a significant amount of information pertaining to the scene being imaged. Automatic image registration relies on detecting and exploiting this common information.

Brown [1992] provides a thorough survey of registration problems, classifying them into four categories: multimodal, template, viewpoint and temporal. This effort deals with the specific problem of multimodal (i.e. multisensor) registration [Manjunath and Fonseca, 1996, Rignot et. al. 1991], which is the registration of images of the same scene acquired from different sensors. Registration methods are characterized in [Brown 92] based on three factors: the geometric transformation model used, the type of image variations the method can handle, and the type of computations performed in order to register the images.

Geometric transformations can be characterized as 2-D or 3-D, and as global or local. When a global transformation model is used, a single set of parameters is computed for the entire image. In the case of a large image, it may be necessary to subdivide it into subregions, each with its own local set of transformation parameters. In this effort, we restrict ourselves to the determination of

global 2-D transformations. The image variations of interest to us are geometric variations---the fact that any given point in the scene may appear at different locations in the two images---and radiometric or photometric variations, which pertain to the fact that this point may have a different pixel value (``intensity'') in the two images.

Different computational approaches can be used for determining the registration parameters once a transformation model is assumed. Methods can be broadly classified as area-based or feature-based. Area-based methods assume that corresponding pixels in the two images will have strongly correlated photometric values. A region in one image can thus be compared to a region in another using some measure of similarity such as cross-correlation. Feature-based methods use information from discrete features such as lines, contours and corner points extracted from the two images.

Most traditional methods for image registration can handle only minor geometric and photometric variations. In a multisensor context, however, the images to be registered in general may be of widely different types, obtained by disparate sensors with different resolutions, noise levels and imaging geometries. The common or ``mutual'' information, which is the basis of automatic image registration, may manifest itself in a very different way in each image. This is because different sensors record different physical phenomena in the scene. For instance, an infra-red (IR) sensor responds to the temperature distribution of the scene, whereas a radar responds to material properties such as dielectric constant, electrical conductivity and surface roughness. Area-based methods are therefore inappropriate for multisensor registration. However, since the underlying 3-D scene giving rise to the shared information is the same, certain qualitative statements can be made about the manner in which information is preserved across multisensor data. Although the pixels corresponding to the same scene region may have different values depending on the sensor, pixel similarity and pixel dissimilarity are usually preserved. In other words, two scene points that have comparable pixel values in one type of image are likely to have comparable pixel values in images

of the scene obtained from other sensors. A region that looks homogeneous to one sensor, is likely to look homogeneous to another, local textural variations apart. Regions that can be clearly distinguished from one another in one image are likely to be distinguishable from one another in other images, irrespective of the sensor used. Although this is not true in all cases, it is generally valid for most types of sensors and scenes. For instance, in SAR, EO and IR images of an area containing a small lake surrounded by bare ground, the perimeter of the lake is likely to be clearly visible, although the photometric appearance of the lake and the surrounding region may be sensor-dependent. Thus features derived from shapes of 2-D image regions contain mutual information that can be used for multisensor registration. In this effort, we approximate region boundaries by polygonal segments, and use these segments along with high curvature points on the boundaries as features for registration.

Man-made objects in a scene also give rise to features that are likely to be preserved in multisensor images. For instance, building edges appear as intensity discontinuities in SAR, EO and IFSAR images, whereas edges of roads appear as discontinuities in SAR and EO images. Feature-based methods that exploit the information contained in region boundaries and in man-made structures are therefore of interest for multisensor registration.

Feature-based methods traditionally rely on establishing feature correspondence between the two images. Such correspondence-based methods first employ feature matching techniques to determine corresponding feature pairs from the two images, and then compute the geometric transformation relating them, typically using a least-squares approach. Their primary advantage is that the transformation parameters can be computed in a single step, and are accurate if the feature matching is reliable. The drawback is that they need to address the highly non-trivial *correspondence problem*: Given 'N' features in each image, the number of possible one-to-one feature mappings is 'N', out of which only one can possibly be correct. Some heuristics can be employed to reduce the number of potential correspondences [Grimson 1981], but this problem

still remains intractable, unless the two images are already approximately registered, or the number of features is small. In [Li and Manjunath 1995], multisensor registration is accomplished by matching contours. This approach will work if the images contain a small number of clean contours. If the data are noisy, however, the contours extracted may be split up into smaller fragments, and contour matching becomes very difficult.

In this effort, we have proposed an approach to multisensor registration that eliminates the need for feature matching. We first decompose the original transformation into a sequence of elementary stages. At each stage, we estimate the value of one transformation parameter by a "feature consensus" mechanism in which all pairs of features are considered as potential matches, and each potential match votes for the value of the parameter that is consistent with it. We introduce the concept of parameter observability to formalize this process.

4.1 THE FEATURE CONSENSUS APPROACH

The basic principle of feature consensus is similar to that used in the Hough transform. The Hough transform maps the feature space into the parameter space, by allowing each feature to vote for a subspace of the parameter space. Clusters of votes in the parameter space are then used to estimate parameter values. Feature consensus is similar in the sense that it involves the determination of vote clusters in parameter space. It is different from the Hough transform in two important aspects: (a) The basic voting unit is a feature pair (f_1, f_2) , where f_1 is from image I_1 and f_2 from image I_2 and (b) The parameter space is one-dimensional, and each voting unit votes for exactly one point in this space. If θ is the parameter being estimated, the features 'f' that vote should possess attributes 'a' that are related by

$$\alpha_2 = g_\theta(\alpha_1) \tag{4.1}$$

where θ is a bijective function that depends only on the parameter being estimated. In other words, θ should be observable with respect to some feature and attribute classes. We assume that the function is of a form that permits the parameter to be written as

$$\theta = h(\alpha_1, \alpha_2) \tag{4.2}$$

The *consensus function* C_θ is defined as

$$C_\theta = \sum_{i,j} \delta\big(h(\alpha_1, \alpha_2)\big), f_i \in F_\infty \tag{4.3}$$

where δ is the discrete impulse function. Feature consensus is simply the process of determining the value of θ that receives the maximum number of votes:

$$\theta_{\max} = \arg\max_{\theta}(C_\theta) \tag{4.4}$$

In order to estimate the transformation parameters a_i by feature consensus, we need to reparametrize the transformation into a set of stages such that at each stage there is a single observable parameter. In mathematical terms, we decompose the original transformation 'T' into a sequence of transformations

$$T_{(a_1, a_1, \ldots a_n)}(I_1) = T^n_{b_n}\left(T^{n-1}_{b_n}\left(\cdots T^1_{b_1}(I_1)\right)\right) \tag{4.5}$$

where $(b_1\ b_2 \cdots b_n)$ are functions of the original transformation parameters $(a_1\ a_2 \cdots a_n)$, of the form

$$b_i : b_i(a_1, a_2, \cdots a_n), i = 1, \cdots n$$

In the simplest case, the 'b' parameters are identical to the 'a' parameters. At each stage 'i', there should be some feature attribute which is transformed in a manner that is dependent only on 'b_i':

$$\exists\ g_{b_i}(\ ) \ni \alpha_2 = g_{b_i}(\alpha_1)$$

The parameters $(b_1\ b_2 \cdots b_n)$ are estimated sequentially by feature consensus, and at each stage 'i' the transformation 'T^i_{bi}' is applied the first image 'I_1', leaving us then with the task of estimating

the remaining parameters '(b_{i+1} $bi+2 \cdots b_n$)' between the transformed first image and the second. This is performed until all the 'b' parameters have been estimated. It is straightforward to estimate the original 'a' parameters of the transformation 'T'. In most cases, this may not be necessary, since the first image can be aligned with the second simply by applying the last stage of the reparametrized transformation.

4.2 AN IMAGE TRANSFORMATION MODEL

A number of choices are available for the global 2-D transformation between two images, as discussed in [Wolberg 1990, Zheng 1993]. The model that is chosen depends on a number of factors, such as prior knowledge about imaging geometries [Chellappa 1996], accuracy required and computational cost. In general, the lowest order model (i.e. the simplest model) should be used. Computational cost increases, and algorithmic robustness decreases, with increase in the complexity of the model. Thus, although a more complex model may be able to capture a wider range of geometric variations, the practical problems associated with model complexity make a simpler (and possibly less accurate) model more attractive.

A decomposition of the from (4.5), satisfying the observability constraint (4.1) may not be obvious for any arbitrary transformation 'T', except in simple cases like the similarity transformation. A similarity transform, which is characterized by four parameters (rotation 'β', translation 't_x', 't_y' and scale 's'), transforms a point '(x,y)' to the point '(X,Y)', according to:

$$\begin{pmatrix} X \\ Y \end{pmatrix} = s \begin{pmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix} \tag{4.6}$$

Using the notation '$p = (x,y)^T$', this transformation can be expressed as a suite of four stages:

$$T(p)=T_{t_y}\left(T_{t_x}\left(T_s\left(T_\beta(p)\right)\right)\right)$$

(4.7)

where

$$T_\beta(p)=\begin{pmatrix}\cos\beta & \sin\beta\\ -\sin\beta & \cos\beta\end{pmatrix}p$$

$$T_s(p)=sp$$

$$T_{t_x}(p)=p+\begin{pmatrix}t_x\\ 0\end{pmatrix}$$

$$T_{t_y}(p)=p+\begin{pmatrix}0\\ t_y\end{pmatrix}$$

In this case, the parameters of the new transformation (β, t_x, t_y, s) are identical to the parameters of the original transformation. The first parameter to be estimated is 'β', the angle of rotation. This parameter is observable from the slopes of line and edge features in the images. If 'l_1' and 'l_2' are corresponding line features with slope angles 'ϕ_1' and 'ϕ_2' respectively, then

$$T_s(p)=sp$$

(4.8)

Thus 'β' can be estimated by consensus of line features, and 'I_1' can be transformed according to (4.8). The scale 's' can be estimated by consensus of pairs of point features, with the distance 'd' between the two points being the candidate attribute:

$$d_2 = sd_1.$$

(4.9)

The translational shifts 't_x, t_y' are observed using point feature location as candidate attribute:

$$p_2 = p_1 + \begin{pmatrix} t_x \\ t_y \end{pmatrix}$$

(4.10)

Let us now consider a more complex quasi-affine model, with five parameters (rotation 'β', translation 't_x', 't_y' and scales 's_x', 's_y'). The transformation can be written as

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{pmatrix} \begin{pmatrix} s_x & 0 \\ 0 & s_y \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix}$$

(4.11)

Proceeding in a fashion similar to that used for the previous case, this transformation can be expressed as a suite of five stages:

$$T(p) = T_{t_y}\left(T_{t_x}\left(T_{sy}\left(T_{sx}\left(T_{\beta}(p)\right)\right)\right)\right)$$

(4.12)

where

$$T_{sx}(p) = \begin{pmatrix} s_x & 0 \\ 0 & 1 \end{pmatrix} p$$

and

$$T_{sy}(p) = \begin{pmatrix} 1 & 0 \\ 0 & s_y \end{pmatrix} p$$

In this case, the rotation 'β' and scales 's_x', 's_y' cannot be observed independently of each other from the slopes and lengths of segments, as in the previous case. The solution to this problem is to find other feature/attribute classes to observe 'β' and scales 's_x', 's_y' independently, or to reparametrize the transformation with respect to a new set of parameters that are all independently observable. However, if the two scale factors are approximately equal (as is true in most cases), the previous method for observing the rotation angle is still valid, and the scale factors can be observed by a simple extension of the previous method.

4.3 NOISE CONSIDERATIONS

The feature consensus scheme may generate a large amount of noise, particularly if the number of features is large. If there are 'N' features in each image, there are a total of 'N^2' votes, of which only 'N' can possibly be correct. Although this is an improvement over the 'N!' complexity associated with correspondence-based methods, the ``signal" component of the consensus function is small compared to the ``noise" component. However, the signal does not get lost in the noise, because the 'N^2-N' incorrect votes will be dispersed over a wide range of values, whereas the 'N' correct votes will (ideally) be clustered around the true parameter value. Thus the mode of the distribution of votes can be identified. For instance, if we are estimating rotation by comparing the slope angles of line features, the incorrect votes will be distributed evenly in the range '($-\pi/2$, $\pi/2$', whereas the correct matches will vote for the true rotation angle 'β'. Nonetheless, it would be helpful to reduce the number of incorrect votes. If, by some simple matching scheme, we can restrict each feature to have a maximum of '$m << N$' possible matches, we can reduce the total number of votes to 'Nm'. For a detailed discussion of robust feature-matching techniques, see [Stewart 1994].

4.4 RESULTS OF THE REGISTRATION ALGORITHM

Our system for estimation of transformation parameters is illustrated schematically in Figure 4.1.

4.4.1 Feature extraction

Robust extraction of features is important for the success of any feature-based multisensor registration scheme. The feature consensus method proposed here relies on the mutual information

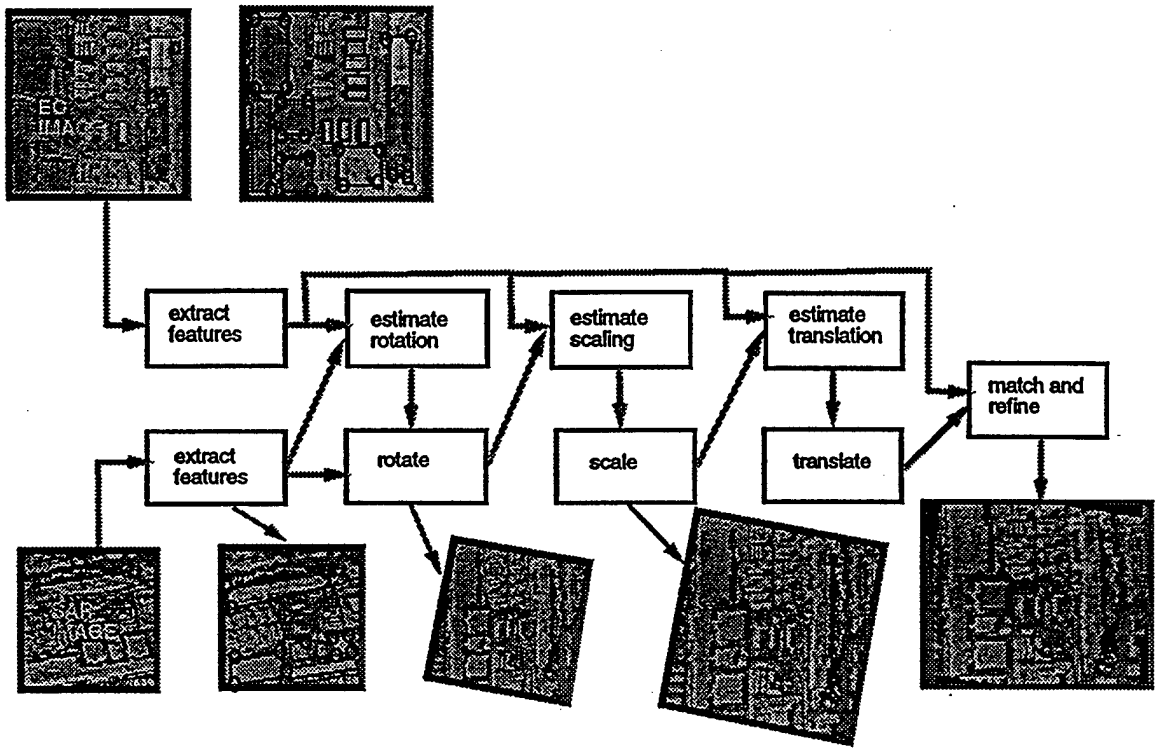

Figure 4.1: A schematic diagram of our feature-based registration technique (applied to Kirkland SAR-EO dataset)

contained in the extracted features in the two images. In other words, a large fraction of the features extracted should be common to both images. This ensures that the votes from matching feature pairs are not lost in the combinatorial noise generated in the feature consensus process. Feature extraction has been one of the most widely researched areas of image processing, and a number of methods have been proposed. In our current implementation, contours are extracted from the images using the Rosin method [Venkatesh and Rosin 1995]. Some post-processing is performed to eliminate noisy contours. Lines are obtained by a polygonal approximation of contours, and then discarding small segments. Feature points are extracted from contours using curvature as criterion. As an example, we show the original SAR and the EO images in Figures 4.2a and 4.2b, and the extracted features are shown in Figures 4.3a and 4.3b.

4.4.2 Estimation of Rotation

Rotation is estimated by consensus of polygonal segments, with slope angle as the attribute. Each pair '(s_1, s_2)' of segments, 's_1' from 'I_1' and 's_2' from 'I_2', produces one vote for the angle of rotation. The vote is weighted by the product of the lengths of the segments. The consensus function for this operation is shown in Figure 4.4. Notice the sharp peak in the consensus function near +100 degrees.

4.4.3 Estimation of Scale Parameters

According to the formulation we explained earlier, two stages of observation are required to estimate the two scale factors 's_x' and 's_y'. These two stages can be combined into a single stage. The candidate feature class consists of pairs of feature points $\langle (x,y), (X,Y) \rangle$ from the same image. The observation equations for the scale factors are simple:

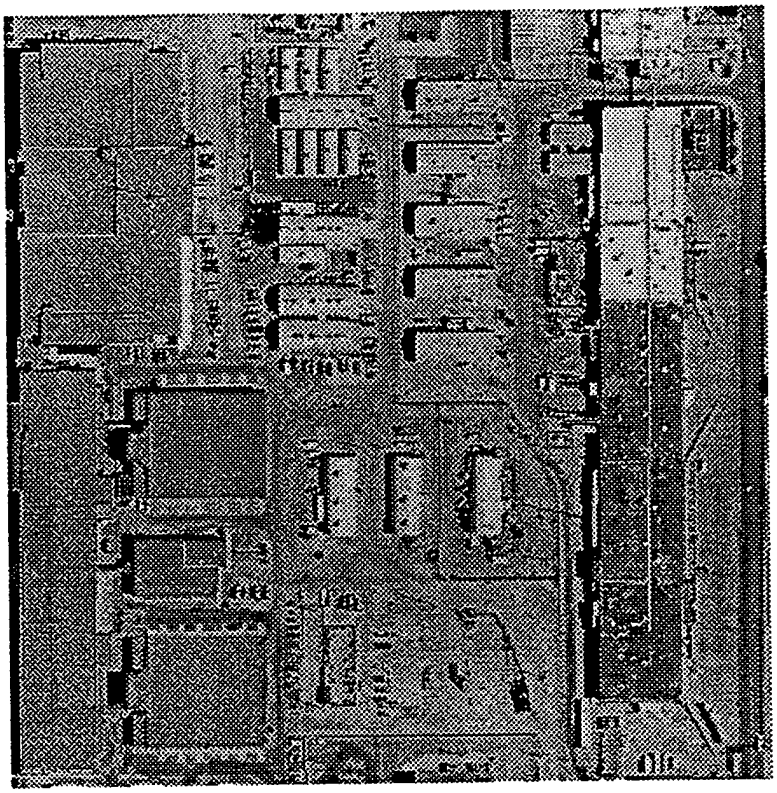

Figure 4.2a: Original EO image of the Kirkland dataset

Figure 4.2b: Original SAR image of the Kirkland dataset

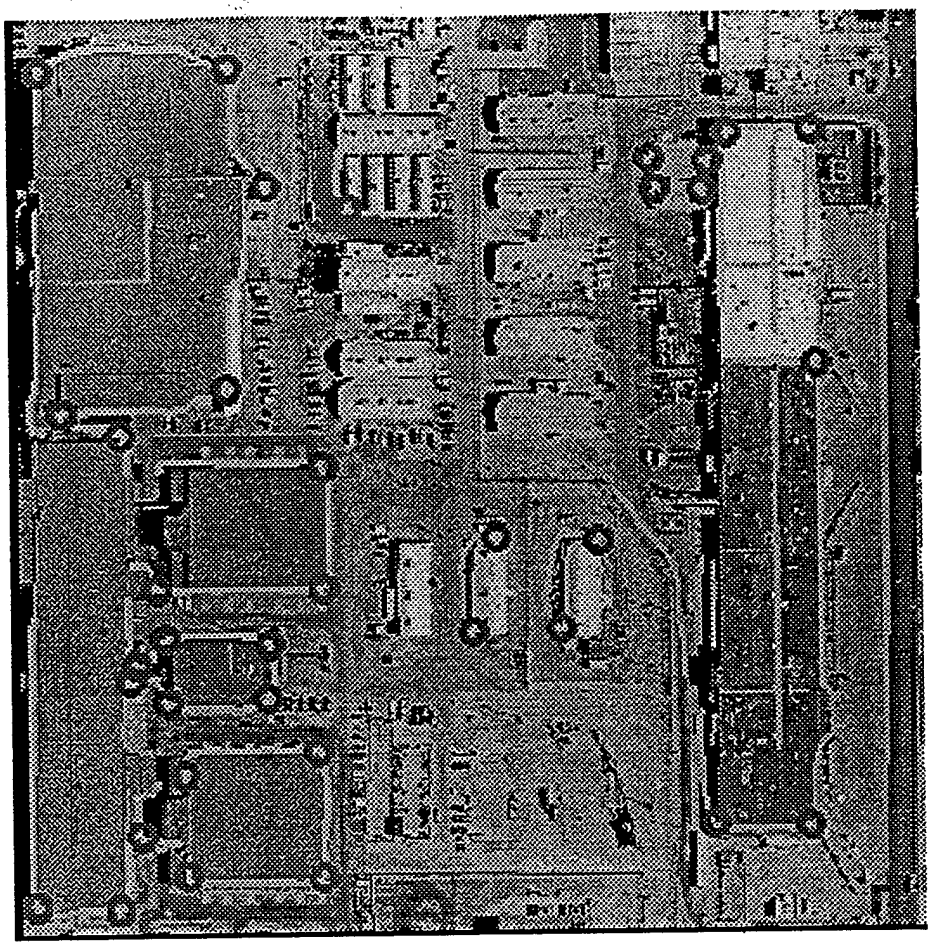

Figure 4.3a: Features of the EO image shown as lines and corners

Figure 4.3b: Features of the SAR image shown as lines and corners

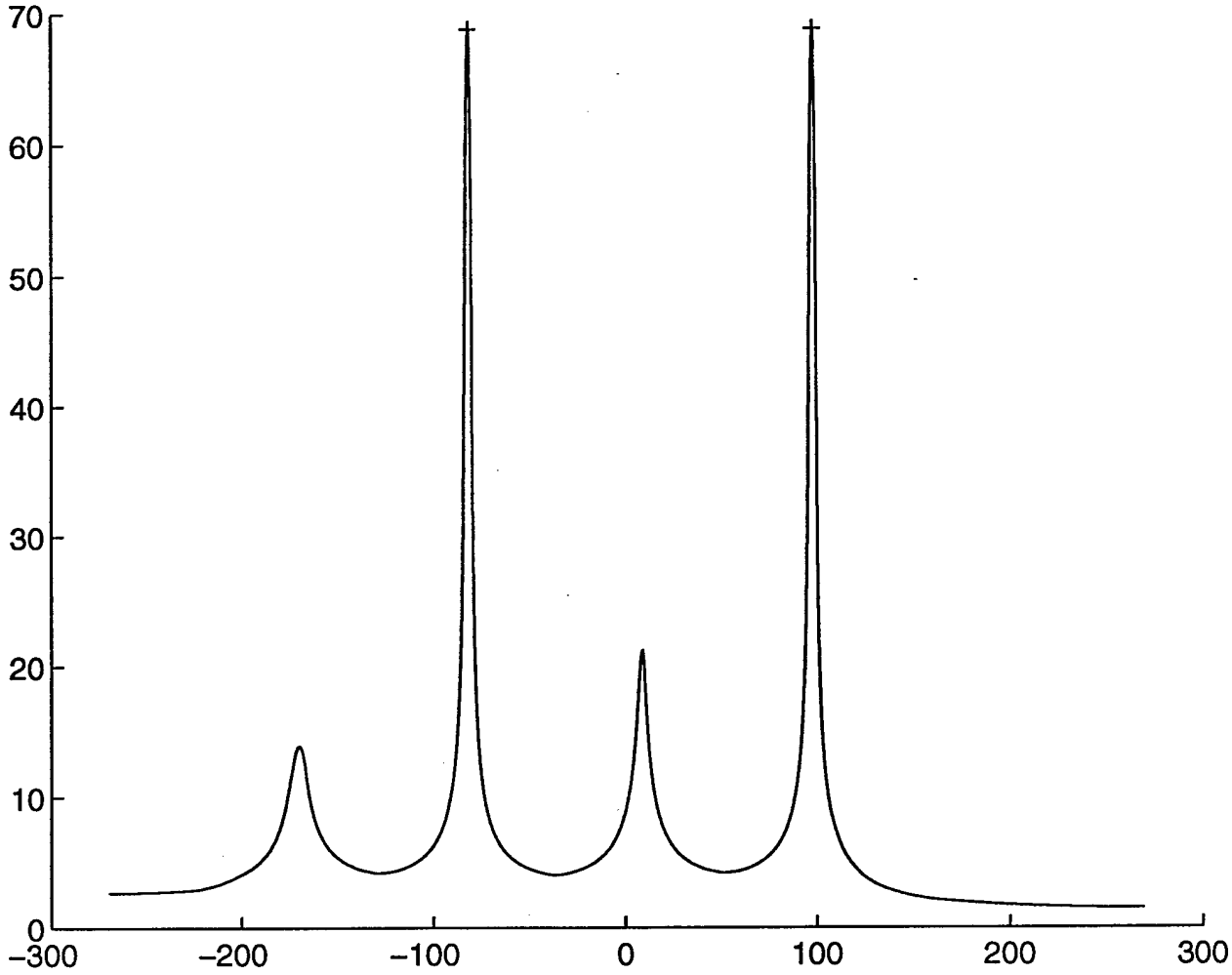

Figure 4.4: Consensus function for the rotation parameter

$$\tilde{x}_2 = s_x \tilde{x}_1$$

$$\tilde{y}_2 = s_y \tilde{y}_1$$

The consensus function for the scale parameters is shown in Figure 4.5.

The 'x' and 'y' scaling factors are approximately 2.1 and 1.6. Whereas the consensus function for 's_x' exhibits an unambiguous peak near the true value, there are two peaks in the consensus function for 's_y'. Currently, the false peak is eliminated by inspection. We need to devise some methods to eliminate outliers to automatically eliminate such false peaks.

4.4.4 Estimation of Translation

Once the rotation and scale have been estimated, estimation of the translation is straightforward. Translational shifts are directly observed from the positions of feature points in the two images. The consensus function is shown in Figure 4.6. The final output of the registration process is shown in the SAR image (see Figure 4.7) with superimposed contours from the EO image. The peaks in the consensus functions for 't_x' and 't_y' are unambiguous. The result of applying this translation is shown by overlaying SAR contours on the EO image.

4.4.5 Refinement

After carrying out the previous stages of processing, the SAR and EO images are approximately aligned. This is sufficient for many applications. However, if further accuracy is required, we can perform a simple nearest-neighbor matching of lines or feature points, and recompute the transformation parameters. Since this is a correspondence-based technique, we can use a more complex transformation model than was used for the feature consensus approach. In our system,

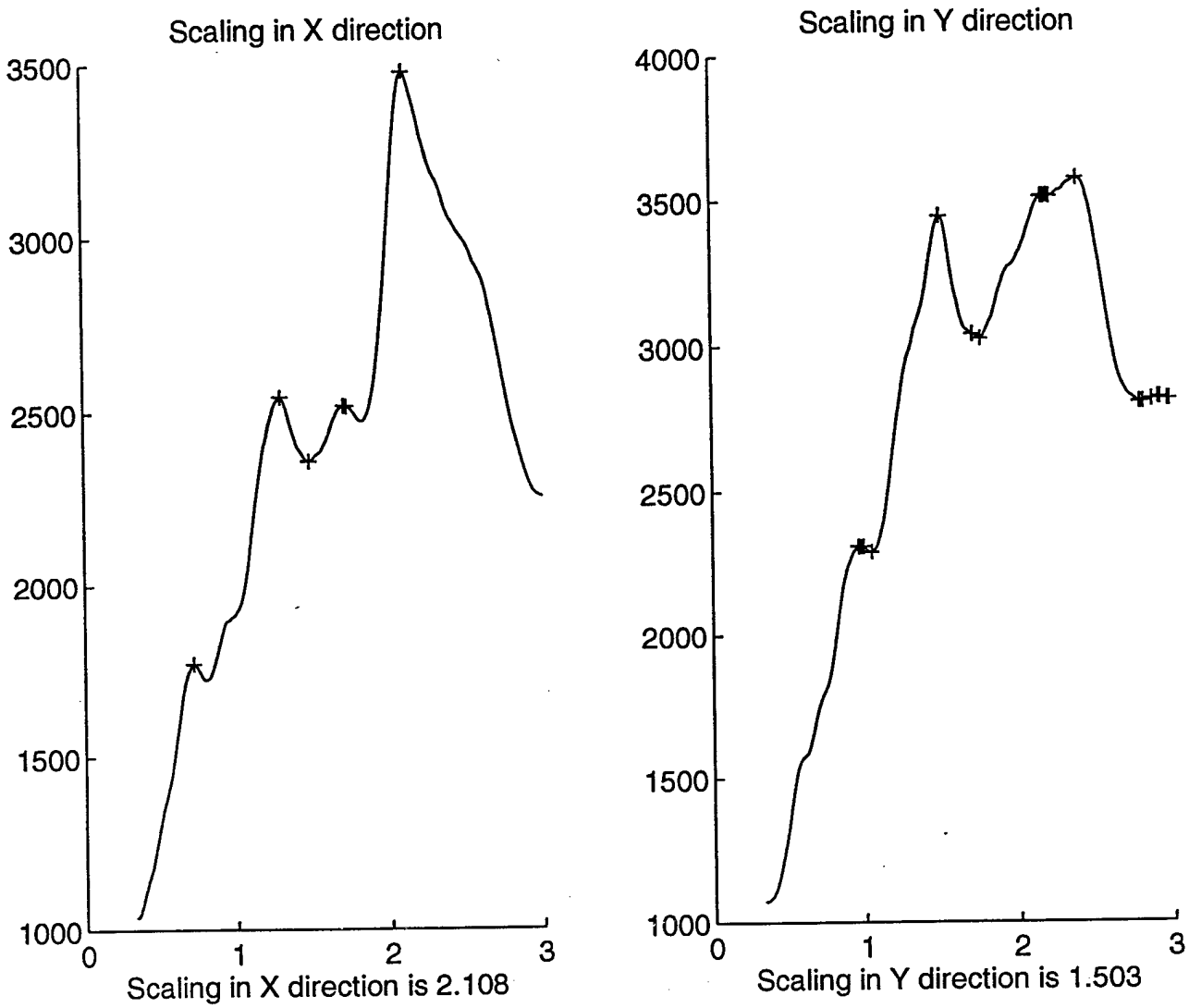

Figure 4.5: Consensus function for the scale parameters

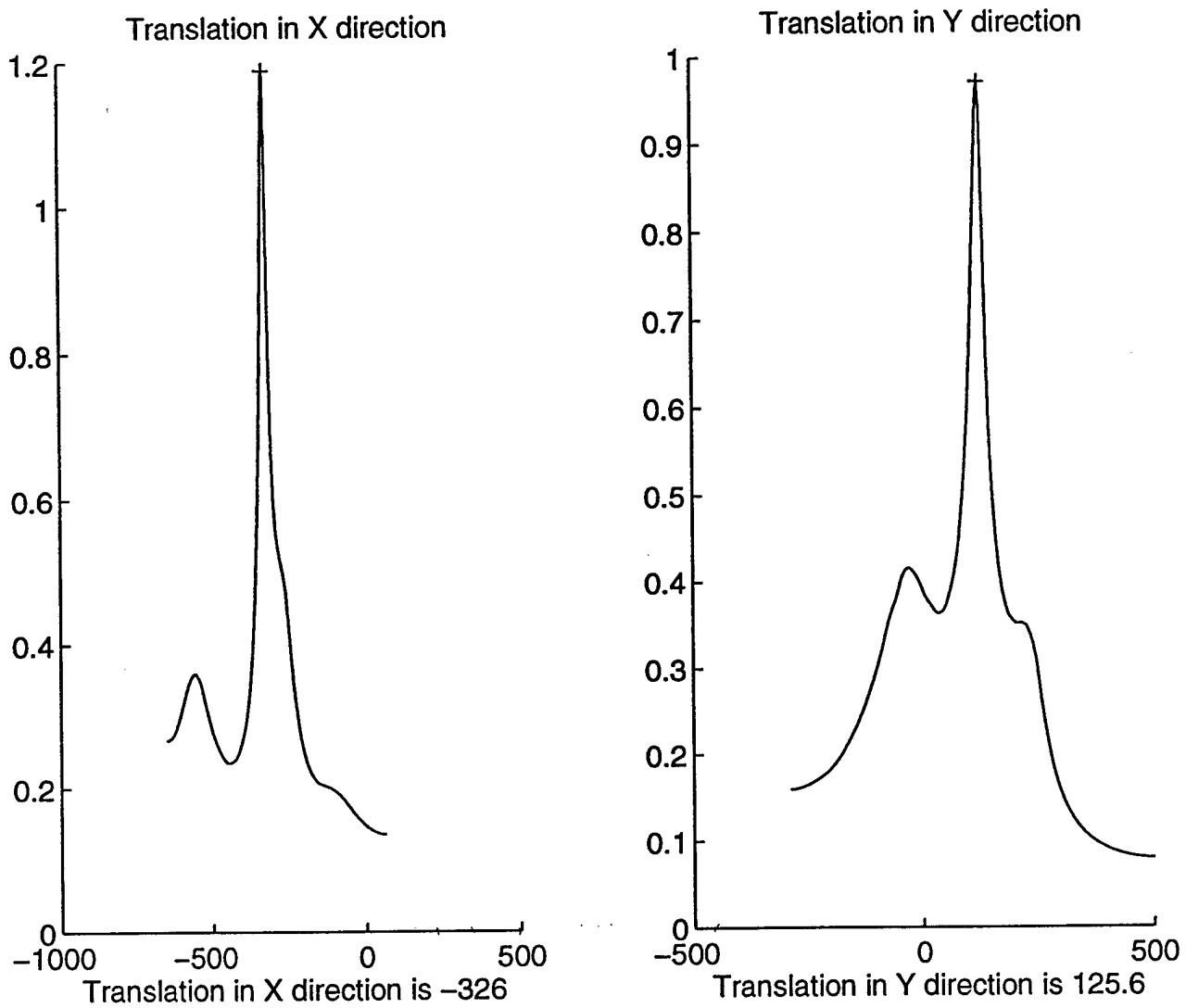

Figure 4.6: Consensus function for the translation parameters

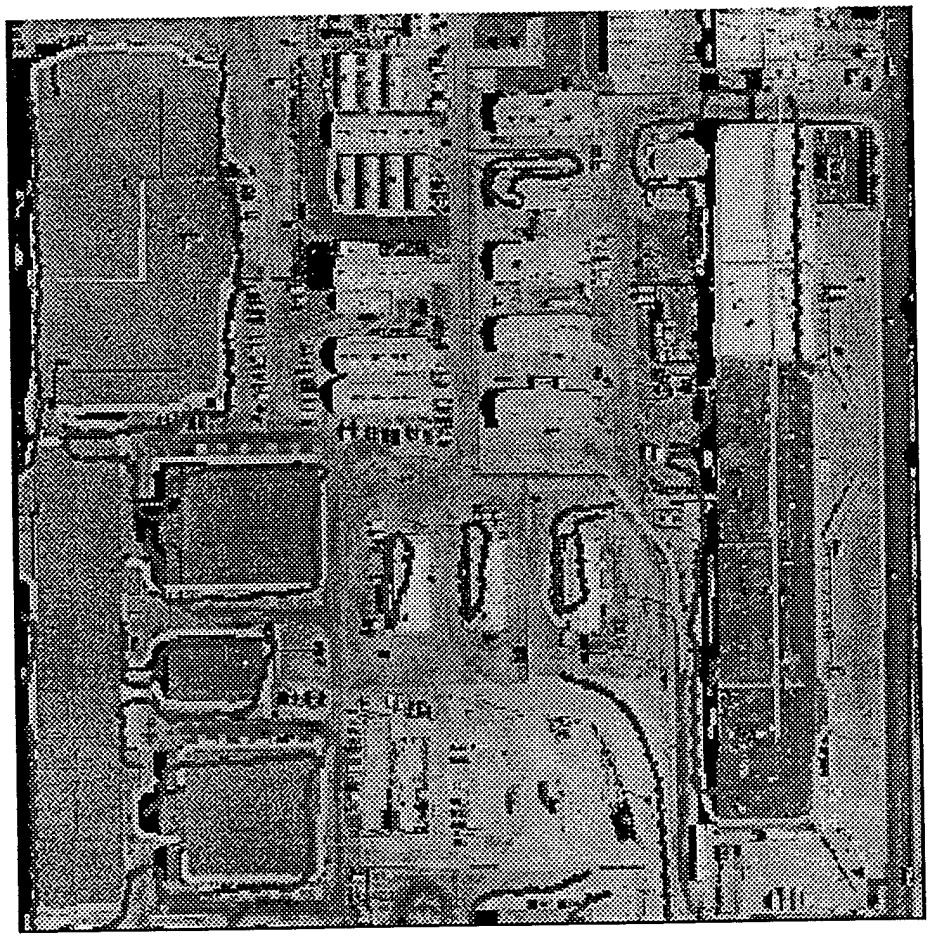

Figure 4.7: Final output of the registration algorithm, super-imposing SAR features on the EO image

we use a projective transformation model for this stage, in which a point '(x,y)' is transformed to '(X,Y)' according to

$$\begin{pmatrix} X \\ Y \\ 1 \end{pmatrix} = c A \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \tag{4.13}$$

where 'c' is an arbitrary scalar, and

$$A = \begin{pmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & 1 \end{pmatrix}$$

For each point correspondence, we can write two equations:

$$X = a_{11}x + a_{21}y + a_{31} - a_{13}xX - a_{23}yX$$
$$Y = a_{12}x + a_{22}y + a_{32} - a_{13}xY - a_{23}yY \tag{4.14}$$

Thus, we get two equations for each point correspondence ' ⟨(x,y), (X,Y)⟩ '. A minimum of four such correspondences are therefore needed to solve for the eight transformation parameters:

$$\begin{bmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -x_1X_1 & -y_1X_1 \\ 0 & 0 & 0 & x_2 & y_2 & 1 & -x_2X_2 & -y_2X_2 \\ \vdots & & & & \vdots & & & \vdots \\ & & & & & & & \end{bmatrix} a = P \tag{4.15}$$

where

$$a = \begin{pmatrix} a_{11} & a_{21} & a_{31} & a_{12} & a_{22} & a_{32} & a_{13} & a_{23} \end{pmatrix}^T$$

and

$$P = \begin{pmatrix} X_1 & Y_1 & X_2 & Y_2 & X_3 & Y_3 & X_4 & Y_4 \end{pmatrix}^T$$

Similarly, given a pair of matching lines 'l = (a b c)^T' and 'L = (A B C)^T', with

$$\begin{pmatrix} a & b & c \end{pmatrix} \begin{pmatrix} x & y & 1 \end{pmatrix}^T = 0$$

$$(A \quad B \quad C)(X \quad Y \quad 1)^T = 0$$

we can write the following two equations:

$$\begin{bmatrix} 0 & Ac & -Ab & 0 & Bc & -Bb & 0 & Cc \\ -Ac & 0 & Aa & -Bc & 0 & Ba & -Cc & 0 \end{bmatrix} a = \begin{pmatrix} Cb \\ -Ca \end{pmatrix} \tag{4.16}$$

Four line correspondences are therefore required to solve for the parameters of the projective transformation.

In our implementation, we use a nearest-neighbor approach to obtain potential point correspondences, and then estimate the transformation parameters using (4.11). Almost invariably, there will be some false matches that will lead to errors in the estimates. A number of methods are available to detect and prune out such ``outliers" [Stewart 1994]. We use the iterative refinement approach developed in [Zheng 1993]. In this approach, transformation parameters are first estimated using the available candidate point correspondences. The computed transformation parameters are used to project points in the first image onto the second. A match '(p_1, p_2)' is considered to be correct if the projection of 'p_1' does not lie too far from 'p_2'. Matches that fail to satisfy this constraint are eliminated, and the transformation parameters are recomputed. This estimate-and-prune step is repeated until all matches satisfy the constraint.

In essence, we have proposed an approach to multisensor registration that does not rely on feature correspondence as its primary mechanism. We argue that feature correspondence cannot be reliably determined in images obtained from disparate sensors, and hence we have proposed a method that is based on feature consensus. By considering all pairs of features as potential matches, and by allowing them to vote for the transformation parameters, we eliminate the need for correspondence. By decomposing the transformation into a sequence of elementary stages, we avoid the complexity associated with Hough transform-style methods. We introduced the notion of parameter

observability to analyze the relationship between features, attributes and transformation parameters. We presented results on real data to validate this approach. Further work in needed on developing better feature detectors for multisensor imagery, and in developing a comprehensive taxonomy of features, attributes and transformations.

5. CFAR DETECTION OF FOA

Many terrain objects, particularly, man-made objects, are fairly strong scatterers. However, the intensity of their radar returns heavily depend on imaging geometry and thus simple thresholding operations will not work. Constant False Alarm Rate (CFAR) processing of SAR data is useful in detecting the metallic objects and other bright reflectors such as the buildings in spatially non-homogeneous clutter. In CFAR processing, the backscatter magnitude at the cell or the region under test is compared to an adaptive threshold, derived from a background threshold window. There are several versions of the CFAR that can be used for this purpose and we describe two of them that were implemented and tested. The methods used are called order statistic (OS) CFAR and cell averaged (CA) CFAR.

The CFAR is essentially a binary hypothesis problem that can be described as the presence or the absence of the object as:

H0 : Target absent (Clutter only)

H1: Target present.

The decision reduces to the likelihood test of the form:

$$\pi_0 \underset{H_0}{\overset{H_1}{\gtrless}} \tau z \tag{5.1}$$

where π_0 is the detector output for the cell under test, z is the statistic derived from the detector outputs from M cells in the background reference window, and τ is the adaptive threshold multiplier. Adaptive threshold τ is related to a given False Alarm rate P_{FA} as :

$$P_{FA} = \int_0^\infty \Pr_0[\pi_0 \geq \tau z] f_z(z) dz \tag{5.2}$$

where $\Pr_0$ is the probability under the null hypothesis .

The K distribution arises when a complex Gaussian process is modulated by a Chi distributed process. Equivalently, it can be shown to arise from modulating the power of a Rayleigh magnitude process with a Gamma distributed variable. This is the case when clutter in a given cell exhibits rapid Rayleigh fluctuations, whose mean is varying slowly over time according to the Gamma distribution. The two parameter K pdf is given by:

$$f_X(x) = \frac{4c}{\Gamma(v)}(cx)^v K_{v-1}(2cx)U(x) \tag{5.3}$$

where v is the shape parameter, $K_v(.)$ is the modified Bessel function of the second kind of order v , and c is a power parameter related to the mean clutter power P_0 by $P_0 = v/c^2$. The shape parameter v controls the "spikiness" of the clutter with the lower value of v implying more spiky clutter. For the K distribution, the False Alarm equation can be written as:

$$P_{FA} = \int_0^\infty \left[\frac{2c^v}{\Gamma(v)}(\tau z)^v K_v(2c\tau z) \right] f_Z(z) dz \tag{5.4}$$

For the OS CFAR processor based on a single ranked sample, the test statistic is the kth ordered statistic from the M reference cells. Hence the probability density function (pdf) of the k th ordered statistic is given by:

$$F_k(u) = k\binom{M}{k} f_0(u) F_0(u)^{k-1} [1 - F_0(u)]^{M-k} \tag{5.5}$$

where $F_0(u)$ and $f_0(u)$ are the univariate clutter cumulative density function (cdf) and pdf under null hypothesis H_0. For the K distribution, the kth order statistic is given by:

$$f_Y(y) = 2ck\binom{M}{k}\left[\frac{2(cy)^{M-k+1}}{\Gamma(v)}\right] K_{v-1}(2cy) K_v^{M-k}(2cy)\left[1 - \frac{2(cy)^v}{\Gamma(v)} K_v(2cy)\right]^{k-1} \tag{5.6}$$

which can be simplified for $v = m + 1/2, m = 0,1,2,.....$ In particular

$$v = 0.5 \Rightarrow f_Y(y) = 2ck\binom{M}{k} \exp[-2c(M-k+1)y][1 - \exp(-2cy)]^{k-1} \tag{5.7}$$

and

$$v = 1.5 \Rightarrow f_Y(y) = 4c^2 k\binom{M}{k} \exp[-2c(M-k+1)y](1+2cy)^{M-k}[1 - (1+2cy)\exp(-2cy)]^{k-1} \tag{5.8}$$

substituting these densities in the false alarm expression leads to an expression for the adaptive threshold that has to be solved numerically.

An example of the CFAR processing is shown on a SAR image of an urban area (see Figures 5.1 and 5.2). The output is binary in that the areas of interest are shown in white. Once the candidate areas of interest are detected in SAR, the corresponding areas in visual or IR imagery are located

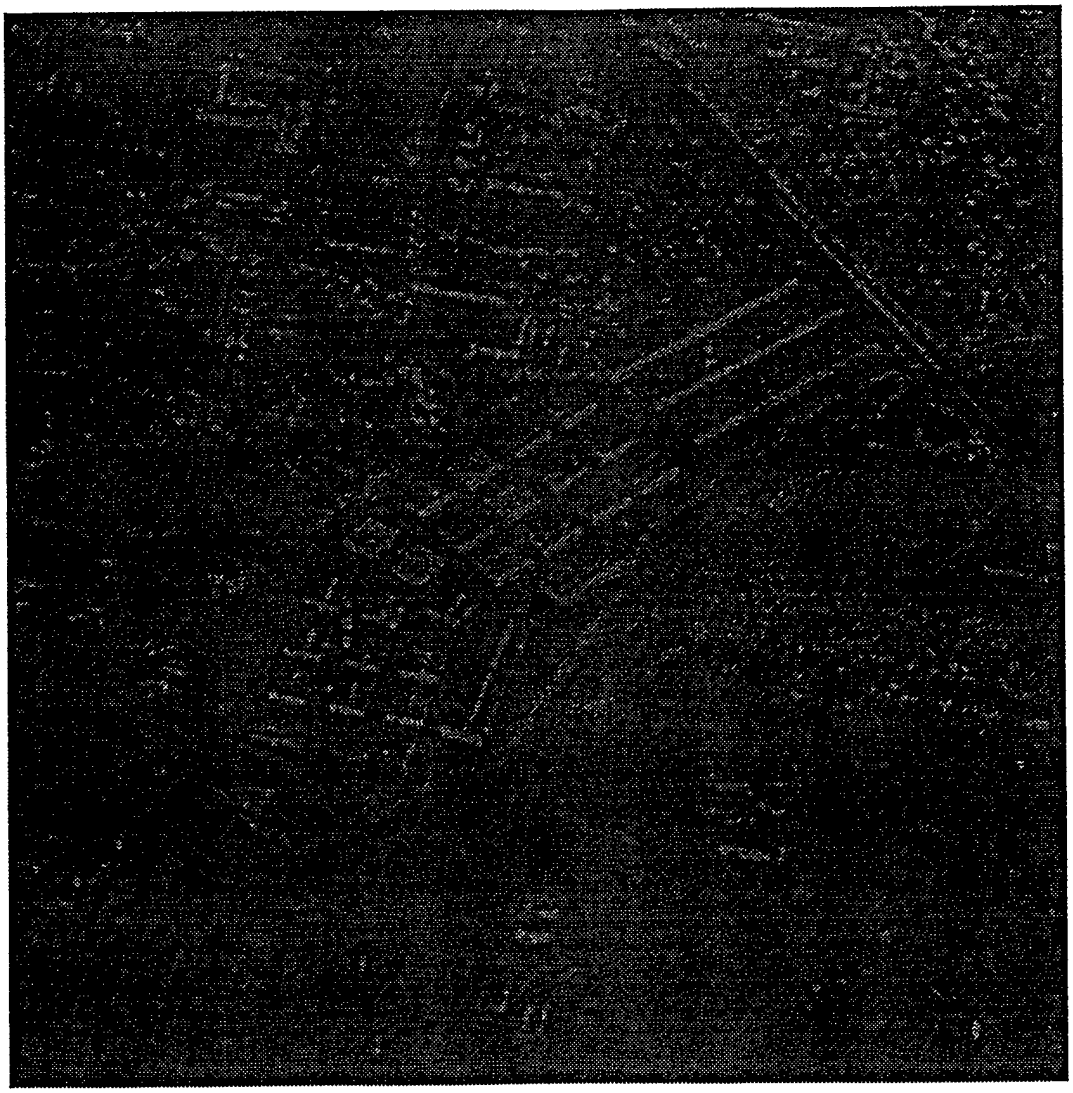

Figure 5.1: Original SAR image of an urban area

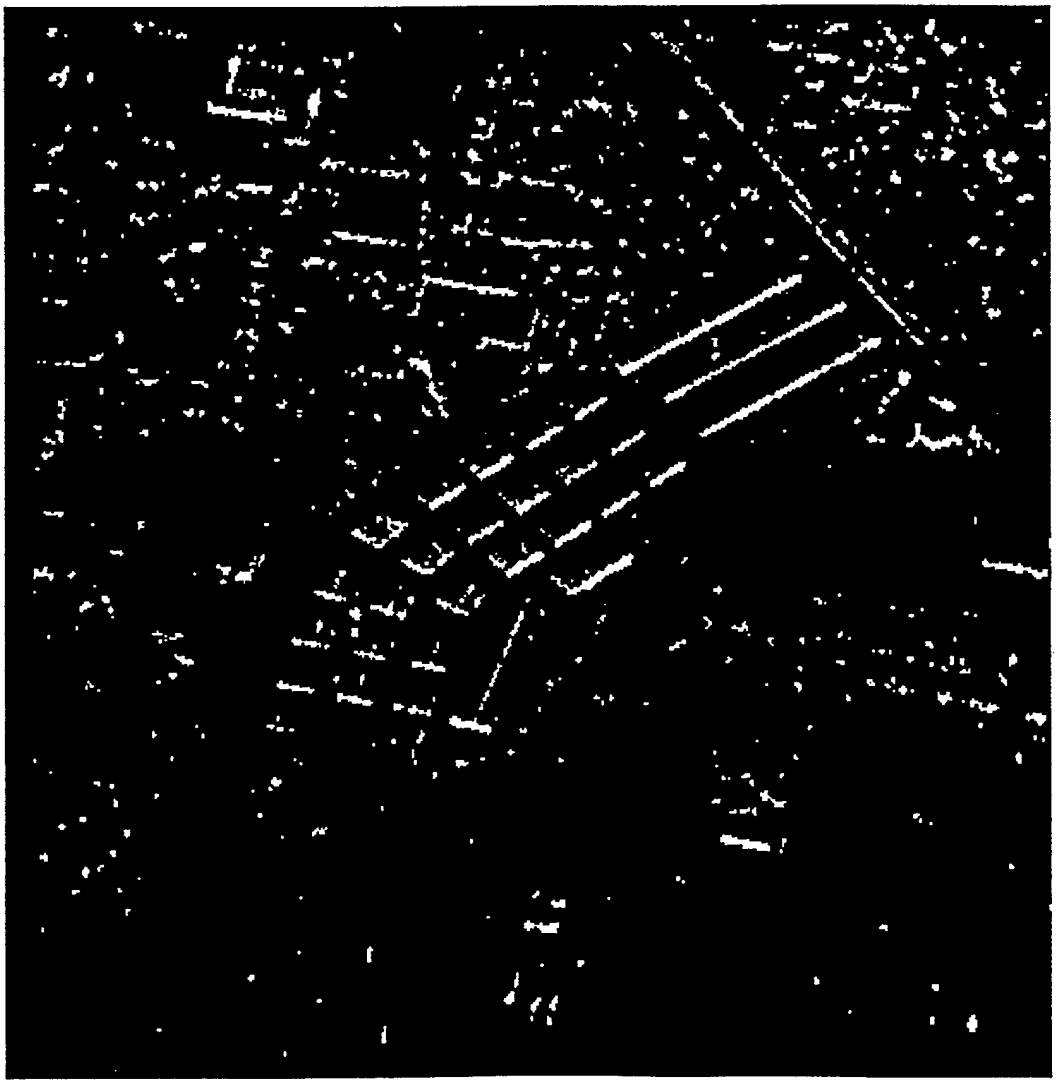

Figure 5.2: Results of CFAR detector output shown as white pixels

by applying the transformation computed using the registration technique explained in Section 4. Unfortunately we have not been able to show the entire process in a single data set due to the lack of co-located data set with good resolution.

6. MODEL BASED RECOGNITION OF TARGETS IN FOA

Unrestricted view point object recognition has continued to be a challenging problem. Model-based vision has provenly demonstrated its usefulness in ATR applications by bringing the complexity of object recognition down to manageable levels by storing *a priori* knowledge of the shape of the objects in a database and matching the model attributes to the image variations. More precisely, the model-based approach involves an object-centered representation of three dimensional models of objects, and a matching scheme to rotate and translate the model to align the model with the image. Popular examples of this approach include ACRONYM [Brooks 1981] and SCERPO [Lowe 1987] (see [Chin and Dyer 1986] for a survey).

An alternative approach to object recognition involves storing normalized multiple views of the object in the model database and finding the "closest" viewpoint appearance to the image instance. This approach is becoming increasingly popular since it mitigates the correspondence and model representation issues. For example, the recent work of Edelman and Weinshall [1989] appears to show that human subjects exhibit better performance in recognizing the objects when the familiarity of multiple views of objects increases. We argue that improved familiarity does not necessarily imply multiple view model storage since familiarity can also aid in generating correspondence hypotheses which in turn result in constrained search of the candidate models. The early perceptual grouping theory of Lowe [1987] is an example of how to achieve constrained search.

Irrespective of convincing biological evidence for either approach, an important computational issue related to multiple-view approach is: how do we *a priori* generate the critical viewpoints or 'aspects' to be stored? Or what constitutes a 'critical view'? Even though the recent work of Seibert and Waxman [1992] discusses this issue and presents a possible solution to this problem via viewpoint clustering, it is inadequate for a real-world problem of arbitrary viewing positions, and missing or additional features other than object silhouette points found in the image. On the other hand, this problem is nonexistent in model-based vision resulting in a major advantage of reduced memory requirements. But, the price we have to pay for this advantage is the additional need to obtain the model-image transformation parameters for matching the candidate models and the corresponding image points. This problem is nontrivial especially when the candidate list is sufficiently large, resulting in a combinatorial explosion.

Our approach to solving this problem involves an efficient multi-level search method. As part of this multi-level search method, we have developed a fast technique to compute the model-image transformation parameters given any three or more corresponding model-image pairs. The motivation of our method comes from the recent work of Huttenlocher and Ullman [1990] who showed that one can obtain the parameters for a unique (up to reflection) model-image transformation under certain restrictions. However, these restrictions are severe enough to limit their wide usage. We have developed two alternatives for their technique, one of which is based on matching feature points, and the other is based matching higher level features, namely, line segments. Before we get into the relevant details, we first would like to provide an overview of our approach.

A schematic diagram of the model-based recognition of targets is shown in Figure 6.1.

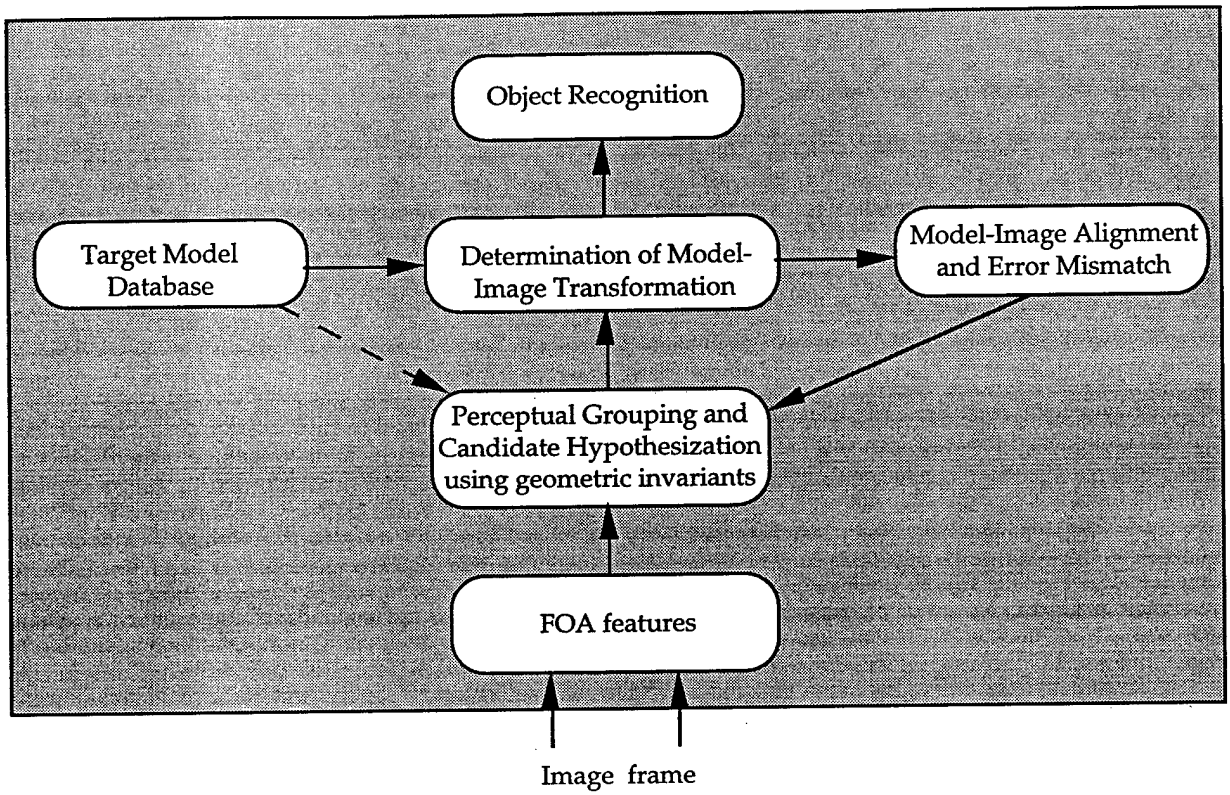

Figure 6.1: The model-based approach to recognition of objects within FOA

The recognition process involves selection of models (indexing) based on geometric invariance followed by the alignment of models with the image. Assuming the resolution of the image sequence is fairly sufficient, we plan to use the BRL CAD models for aligning the models with the image. We have developed two different techniques to do this alignment, one of which is point based and the other is line based. We begin with the point based formulation.

6.1 POINT-BASED FORMULATION

The image projection is assumed to be the scaled-orthographic. That is, we approximate the perspective projection by the orthographic projection with a scaling factor. Thus we have,

$$\begin{bmatrix} X \\ Y \end{bmatrix} = s \begin{bmatrix} x \\ y \end{bmatrix}$$

(6.1)

where (X, Y) is a point in the image which corresponds to the point (x, y, z) in the model and s is the scale factor. This approximation is valid for most real-world applications especially in the case of air-ground target acquisition where $\Delta z/z$ is expected to be small when compared to z. We follow the convention that when we specify a quantity in the object-centered frame of reference we subscript the quantity with 'o'. For a point $(x_o\ y_o\ z_o)$ in the object-centered frame of reference, the corresponding point projected in the image plane according to the transformation, referred to as the π-transformation, is given by:

$$\begin{bmatrix} X \\ Y \end{bmatrix} = s \left[\begin{bmatrix} T_x \\ T_y \end{bmatrix} + R_c \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} \right]$$

(6.2)

where, R_c contains the first two rows of the rotation matrix R. Typically R, in the roll-yaw-pitch formulation with (μ_x, μ_y, μ_z) as the rotation angles. It is clear from (4) that the π-transformation involves six parameters $(s, T_x, T_y, \mu_x, \mu_y, \mu_z)$. Intuitively, it follows that to solve for the six parameters, we require at least three corresponding model-image pairs which generate six error quantities between projected image coordinates of the model points and actual image points.

Our algorithm to compute the model-image transformation is a Newton's method in that it involves computing a set of error quantities which are then used to drive a nonlinear system towards its solution in the direction of minimizing these error quantities. More precisely, if a solution for the nonlinear system of variables 'U' (i.e. $U = [s, T_x, T_y, \mu_x, \mu_y, \mu_z]^T$) is required, Newton's method allows the following iterative error correction such that

$$U^{i+1} = U^i - c,$$

(6.3)

where the required correction c is related to the model-image error 'e' through the Jacobian 'J' of partial derivatives ($\partial e_i/\partial c_j$) as follows. That is,

$$Jc = e. \tag{6.4}$$

The error vector used to correct the current estimates is simply:

$$e = \begin{bmatrix} X \\ Y \end{bmatrix} - s\left[\begin{bmatrix} T_x \\ T_y \end{bmatrix} + R_c \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix}\right] \tag{6.5}$$

The correction vector c can then be obtained by solving the linear system in (6.4). Note that R_C contains only the first two rows of the rotation matrix R and thus, it is a 2 X 3 matrix. The Jacobian 'J' contains two columns corresponding to only one of the three pairs of the model-image tuples. It may be noted that to solve the linear system of six unknowns in the correction vector 'c' we need to obtain the complete Jacobian by adding the columns due to the remaining two corresponding pairs. A major advantage of the proposed iterative technique is its fast convergence. Now, in order to extend this approach to a least-squares formulation, we only need to add two additional columns to J^T and two additional rows to 'e' for every additional model-image pair. However, since there are only six unknowns the system becomes over-constrained. By pre-multiplying (6.4) by J^T we get:

$$J^T Jc = J^T e \tag{6.6}$$

We have implemented this algorithm for a sequential machine and have found that the convergence occurs rapidly within a few iterations.

6.2 LINE-BASED CLOSED-FORM FORMULATION

The point based formulation has a major drawback reliable extraction of feature points. Often, feature points are difficult to extract from noisy images. In such cases, lines are easier to extract as long as we do not attach any significance to the end-points of these lines. That is, we treat each line segment as though it is an infinite line with a certain intercept and slope. This way, the line based formulation is fairly robust.

Let $P_i = \begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix}$ and $P_j = \begin{bmatrix} X_j \\ Y_j \\ Z_j \end{bmatrix}$ be two end points in 3-D which define the line segment of interest to us. The projections of these points are related to their 3-D counterparts as:

$$p_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix} = s \begin{bmatrix} T_x \\ T_y \end{bmatrix} + s \begin{bmatrix} < R_x, P_i > \\ < R_y, P_i > \end{bmatrix} \tag{6.7}$$

For each model line L_k with its end points $\left(P_{k_1}, P_{k_2}\right)$ in 3-space and corresponding end points $\left(p_{k_1}, p_{k_2}\right)$ in the 2-d image we can show that

$$\left(p_{k_1} - p_{k_2}\right) = \begin{pmatrix} x_{k_1} - x_{k_2} \\ y_{k_1} - y_{k_2} \end{pmatrix} = s \begin{pmatrix} \langle R_x, P_{k_1} - P_{k_2} \rangle \\ \langle R_y, P_{k_1} - P_{k_2} \rangle \end{pmatrix} \tag{6.8}$$

After some algebra involving this equation for the two end points of the three lines, we can show that:

$$AR_x = BR_y \tag{6.9}$$

where the 3X3 matrices A and B are given by:

$$A=\begin{bmatrix}(y_1^1-y_2^1)(P_1^1-P_2^1)^T\\(y_1^2-y_2^2)(P_1^2-P_2^2)^T\\(y_1^3-y_2^3)(P_1^3-P_2^3)^T\end{bmatrix} \text{ and } B=\begin{bmatrix}(x_1^1-x_2^1)(P_1^1-P_2^1)^T\\(x_1^2-x_2^2)(P_1^2-P_2^2)^T\\(x_1^3-x_2^3)(P_1^3-P_2^3)^T\end{bmatrix}$$

Assuming B is invertable, we can introduce 'W' such that $W = B^{-1}A$. Now, by imposing the orthogonality and the unit-length constraints, we get:

$$R_x^T R_x = R_x^T W^T W R_x = 1 \tag{6.10}$$

$$R_x^T W^T R_x = 0 \tag{6.11}$$

It may be noted that W contains all the known quantities relevant to the three model line image pairs in consideration. The solution to (6.10) and (6.11) can be reduced to a solution for a fourth order polynomial. To demonstrate this we first perform an SVD on W,

$$W=U\Lambda V=\begin{pmatrix}u_1 & u_2 & u_3\end{pmatrix}\begin{pmatrix}\lambda_1 & 0 & 0\\0 & \lambda_2 & 0\\0 & 0 & \lambda_3\end{pmatrix}\begin{pmatrix}v_1 & v_2 & v_3\end{pmatrix}^T$$

Since V is the SVD is also an orthonormal basis of R^3, we can write $R_x = av_1 + bv_2 + cv_3$. With

$$w_{ij}=\langle v_i, u_j\rangle$$

we get the system:

$$\begin{aligned}&a^2+b^2+c^2=1\\&\lambda_1 a^2+\lambda_2 b^2+\lambda_3 c^2=1\\&\lambda_1 w_{11}a^2+\lambda_2 w_{22}b^2+\lambda_3 w_{33}c^2+(\lambda_1 w_{21}+\lambda_2 w_{12})ab+(\lambda_2 w_{32}+\lambda_3 w_{23})bc+(\lambda_2 w_{13}+\lambda_3 w_{31})ac=0\end{aligned} \tag{6.12}$$

With some tedious algebra, we can show that the rotation parameters involving the rotation matrix R reduce to a 4th order polynomial:

$$l_1\alpha^4 + l_2\alpha^3 + l_3\alpha^2 + l_4\alpha + l_5 = 0 \tag{6.13}$$

After we solve for the rotation parameters from the above equation, we can back-substitute in the projection equation to get the scale and translation parameters. We have avoided some tedious math and substitutions for clarity.

The above quartic equation will have zero, two or four non-negative roots that can give rise to between zero and 32 real-valued solutions to the system (6.12 with back substitutions) in the general case. The true solution will always emerge in the zero noise, correct correspondence case. For each non-negative solution α to the quartic, there is upto one non-negative value of β and upto one non-negative value of c^2 that can be computed from the first two equations of the system (6.12). Since each valid set (a^2, b^2, c^2) gives rise to 8 possible alternatives ($\pm a, \pm b, \pm c$), and there are upto four such sets, there are potentially 32 solutions. Validity of each of the solutions is verified by evaluating the third equation of (6.12). Once the values of (a,b,c) are known, the rotation vectors R_x, and R_y can be determined. The scale factor 's' can be computed from (6.8). Also the translations T_x and T_y can be recovered from (6.7) after substituting for the scale and the rotation quantities.

6.3 SOLVING THE CORRESPONDENCE PROBLEM

It may be noted that both the line based and the point-based alignment methods help solve the problem of pose estimation give the sets of corresponding model-image feature pairs. We still have not solved the model-image correspondence problem. The model-image correspondence problem is a very challenging and complex problem by itself, requiring a lot more attention than a

Phase I effort can focus. However, we have investigated the underlying complexity and developed an interim solution which works well for a limited model set.

We begin explaining this method by making an important observation as follows. When there are N model and M(<N) image lines, there are $\left(\frac{N!}{(N-1)!}\right)$ permutations of model to image lines. However, not all these permutations lead to the correct solution(s). There are several simple heuristics that we can employ to solve this combinatorial problem. For instance, on the basis of adjacency, the permutation can be reduced to a much smaller quantity. This is performed by mapping triplets of adjacent image lines with triplets of model lines having a common vertex. The search space can be substantially reduced this way. We can also compare the ratios of lengths of image lines meeting at an intersection with those of the hypothesized model pairs and prune out the impossible pairs. There are several other such heuristics which can be used to reduce the search space.

Once the search space is reduced, we initiate a following clustering technique (similar to a Hough transform) on the constrained search space involving all the normalized rotation angles (r_x, r_y, r_z) as follows.

1 Initially, the entire "cloud" of points is considered to lie on a single cluster. The size of the cluster is set to 1.0 (since the normalized rotation angles lie within the unit cube.

2 At each iteration, the center of mass of the current cluster is computed. The cluster size is shrunk by a constant factor (usually a number close to 1.0, say 0.95).

3 The iterations are repeated till the cluster size shrinks to a small, predetermined value (say 0.05). The centroid of the resulting cluster is chosen to be the cluster center.

Using the above combinatorial process for guessing matching triplets, followed by a closed form solution to the guess and clustering on all solutions obtained gives the cluster center. Often, we have experimentally observed to correspond to the true solution. To introduce additional noise robustness, we can then use these initial guess and all the high confidence model-image feature pairs in a non-linear least-squares method.

In the Phase II effort for final prototype development, we will replace strengthen this approach with additional heuristics. Also, the model selection process will be improved by using geometric invariants [Weinshall 1993, Weiss 1993] and simple cues such as aspect ratios of targets for heuristics based ranking. We will also investigate other major issues such as the alignment metric.

6.4 RESULTS OF MODEL-BASED ATR

We show the results of our model-based ATR module on the MUSTRS data set. The original FLIR image of the zil truck with the lines extracted is shown on Figure 6.2. The clusters of closed form solutions visualized from different perspectives are shown in Figure 6.3a and 6.3b. The "bunching" of the solutions can be clearly seen at the top of the image. Using the cluster center, the rotation angles, scale and translation were obtained and these were passed to the non-linear least-squares module. The recognized 3-D model of the zil-truck at the end of the model-based alignment process shown in Figure 6.4. The mirror as well as true solutions are shown in Figure 6.5. The mirror solution can be removed only if we consider more sophisticated alignment

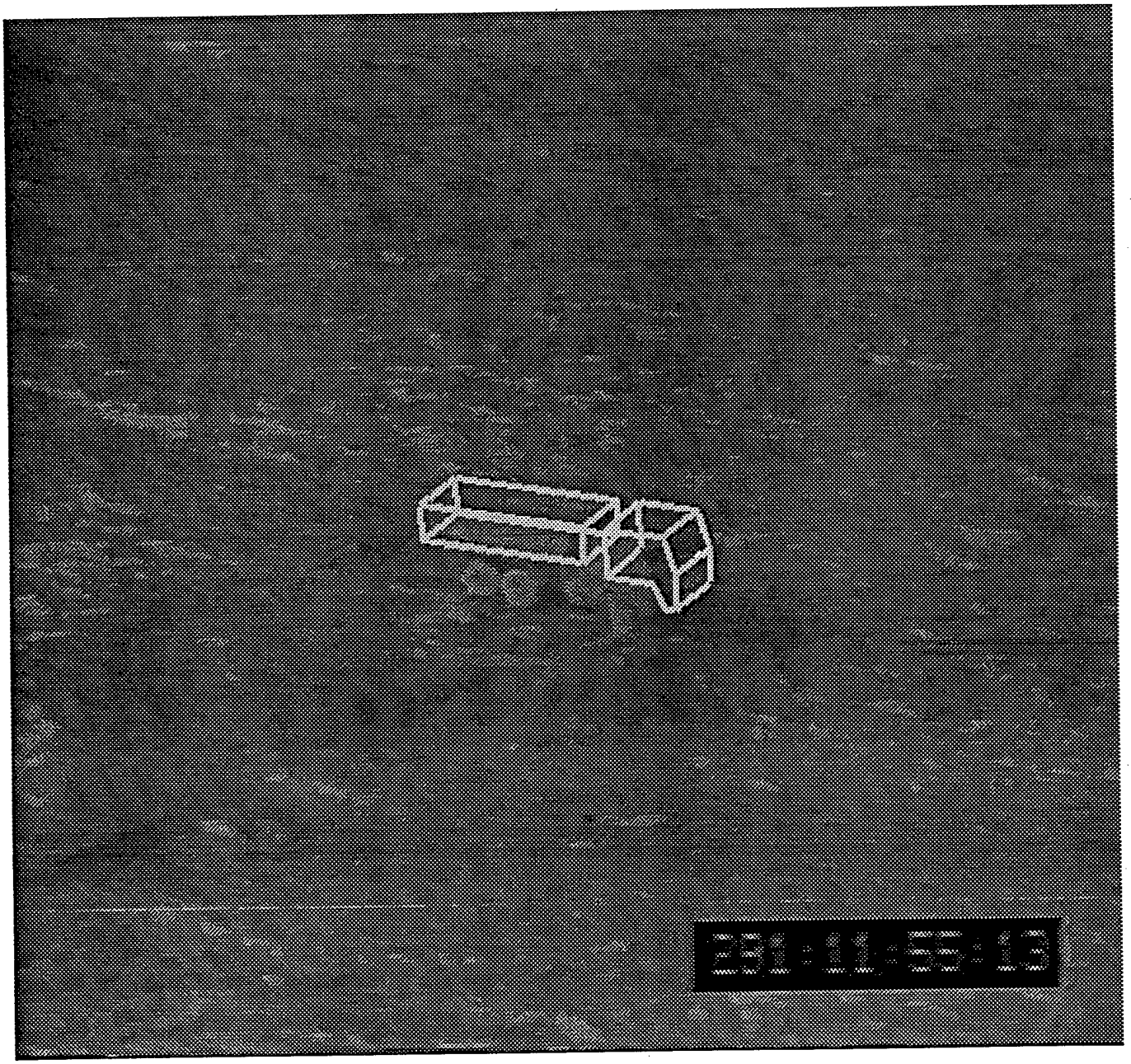

Figure 6.2: Lines extracted on a FLIR image of the MUSTRS data set

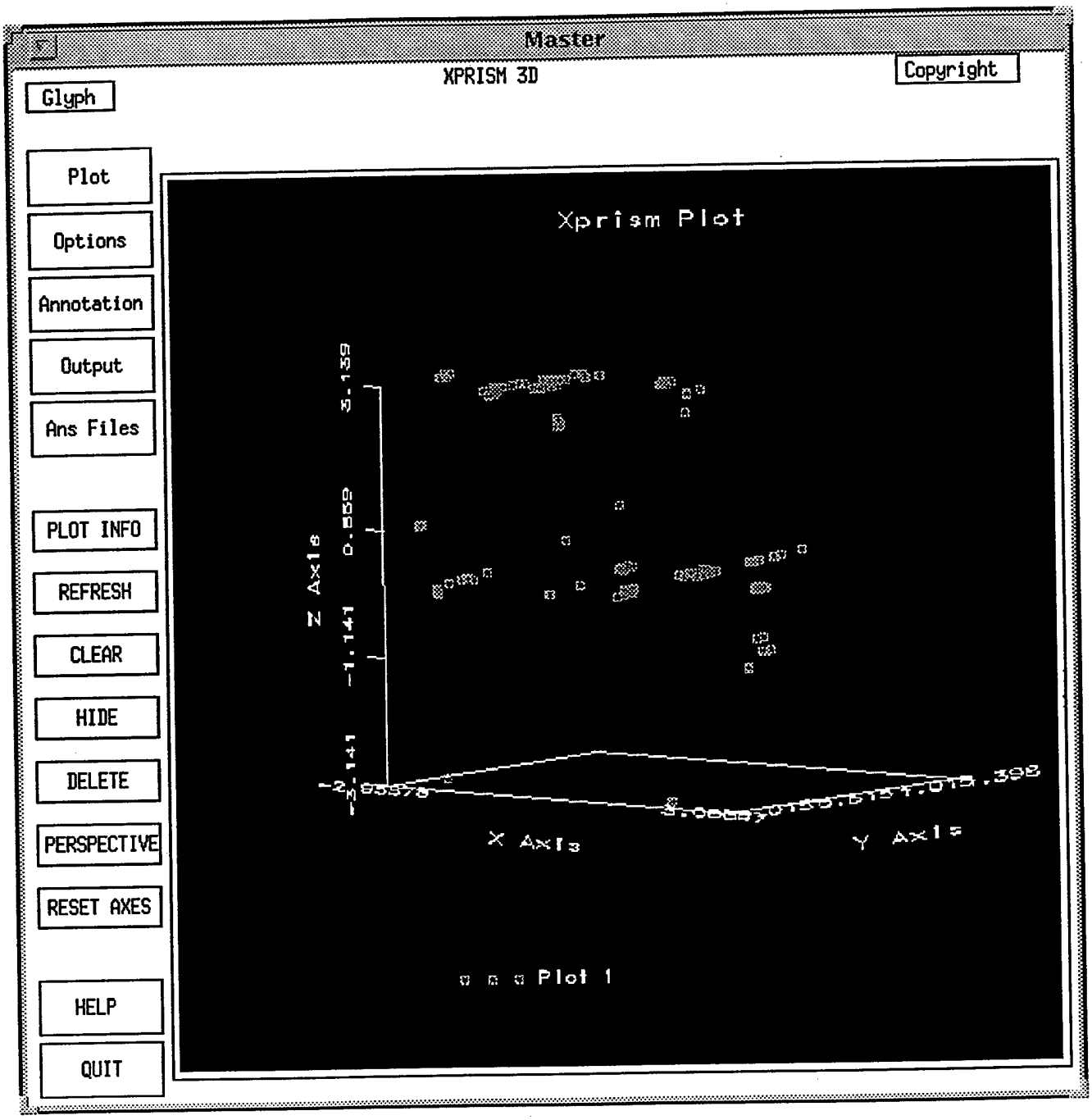

Figure 6.3a: Visualization of the clusters in the parameter space

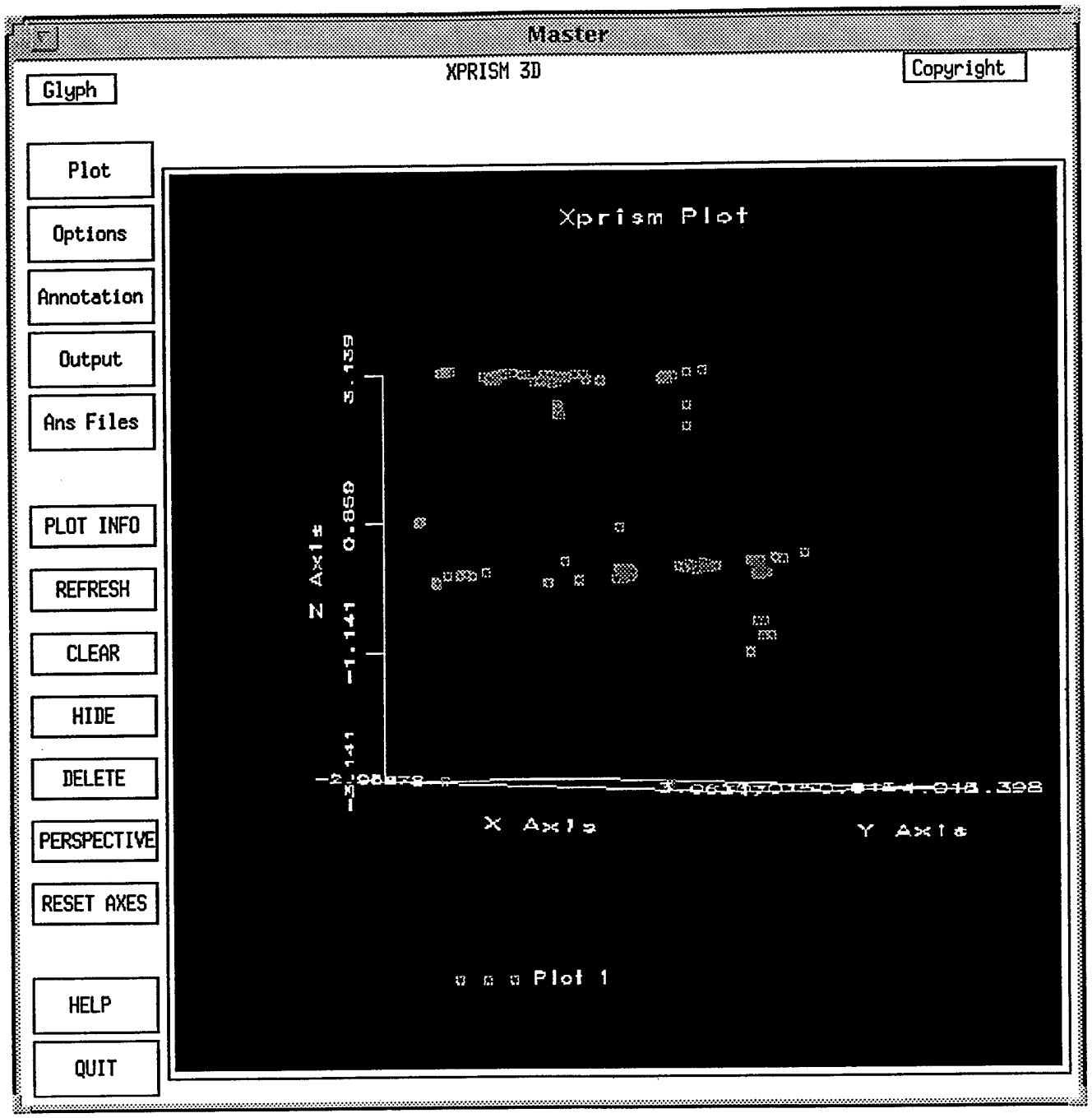

Figure 6.3b: Visualization of the clusters in the parameter space

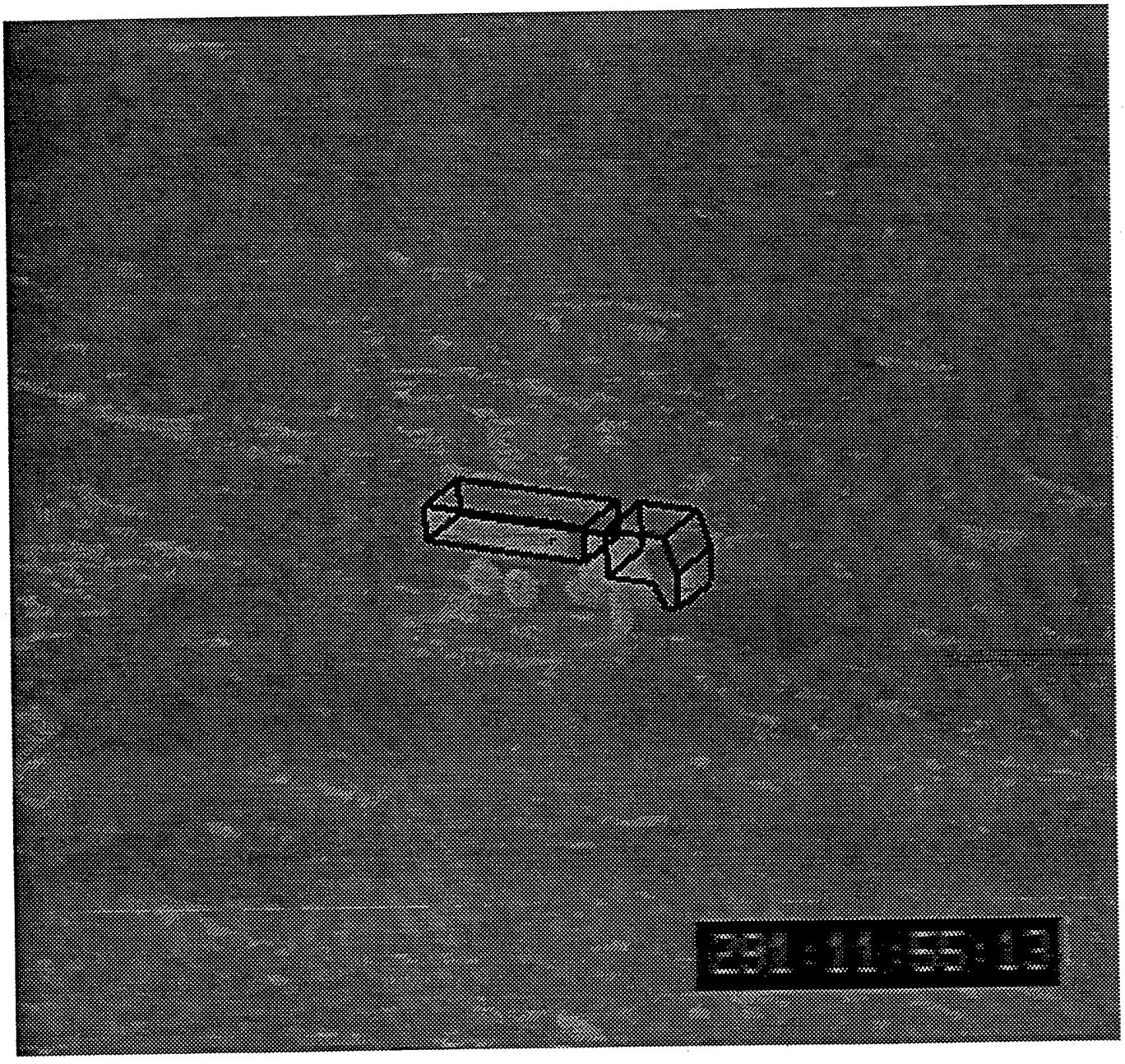

Figure 6.4: Results of aligning the wire-frame model on the IR image

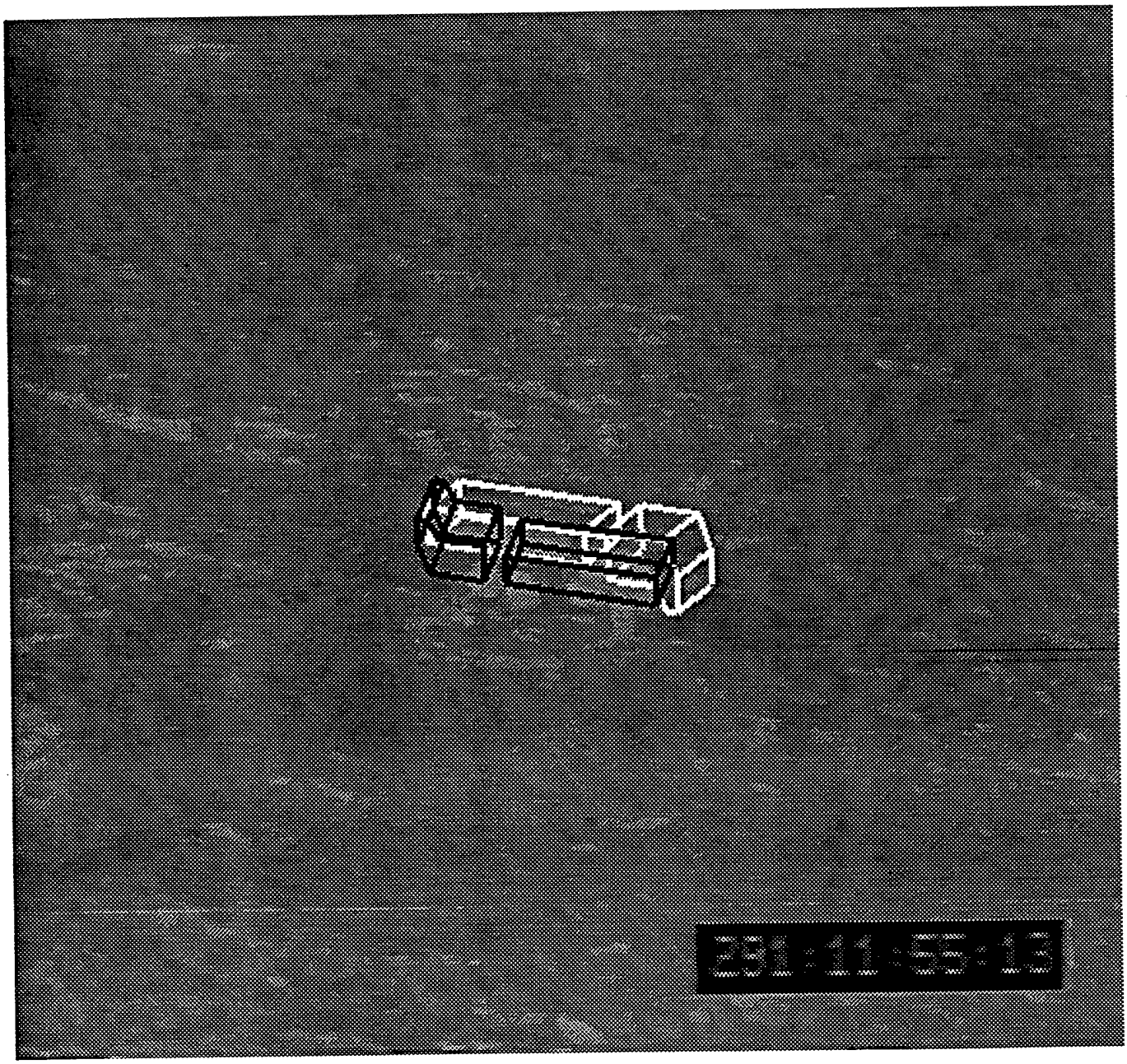

Figure 6.5: The correct solution and its mirror reversal aligned on the IR image of the zil-truck

metrics. With the Euclidean measure, both the solutions will have strong peaking in the parameter space.

7 REAL-TIME FEASIBILITY

To demonstrate the real-time feasibility of the algorithms developed in this STTR, we have implemented the CFAR techniques on an embedded board. Since the computations are locally parallel for these algorithms, we can make use of a simple Single Instruction Multiple Data (SIMD) type hardware to compute most operations in parallel. One such hardware is the CNAPS which is quite inexpensive and commercially available as PCI and ISA boards for about the price of a PC.

The CNAPS system is a parallel hardware that can be configured as a dedicated Ethernet resource for Sunsparcstations and Hewlett Packard workstations with an architecture optimized for executing image processing, pattern recognition and neural network algorithms. Figure 7.1 illustrates the CNAPS architecture. This system is a linear, single-instruction-multiple-data (SIMD) array of state of the art custom CMOS processors each equipped with a block of local memory.

A CNAPS sequencer supervises the execution of instructions on these processors. The sequencer broadcasts parallel input data to every processor in the array on an input bus, and broadcasts instructions on a command bus. The processors execute the instructions simultaneously on the data stored in their local memories.

The program and file memory areas store the assembled machine code and segments of data respectively. A 32-bit internal bus forms the interconnection network between the sequencer, program and file memory areas and the external interface.

The currently available CNAPS processors have a clock speed of 20 MHz. The largest of these configurations has 512 processors, giving a throughput of approximately 10 billion instructions per second. The complex multiply-accumulate instruction desired for neural network operations can be performed in one clock cycle as the internal circuitry on the sequencer splits the clock into four phases. The local memory on each processor is currently 4096 byte, and the global data memory (file memory) is 32 Mbyte. In addition, the Motorola 68030 based controller that interfaces between the host and the server provides 4 Mbyte of DRAM for storing control software.

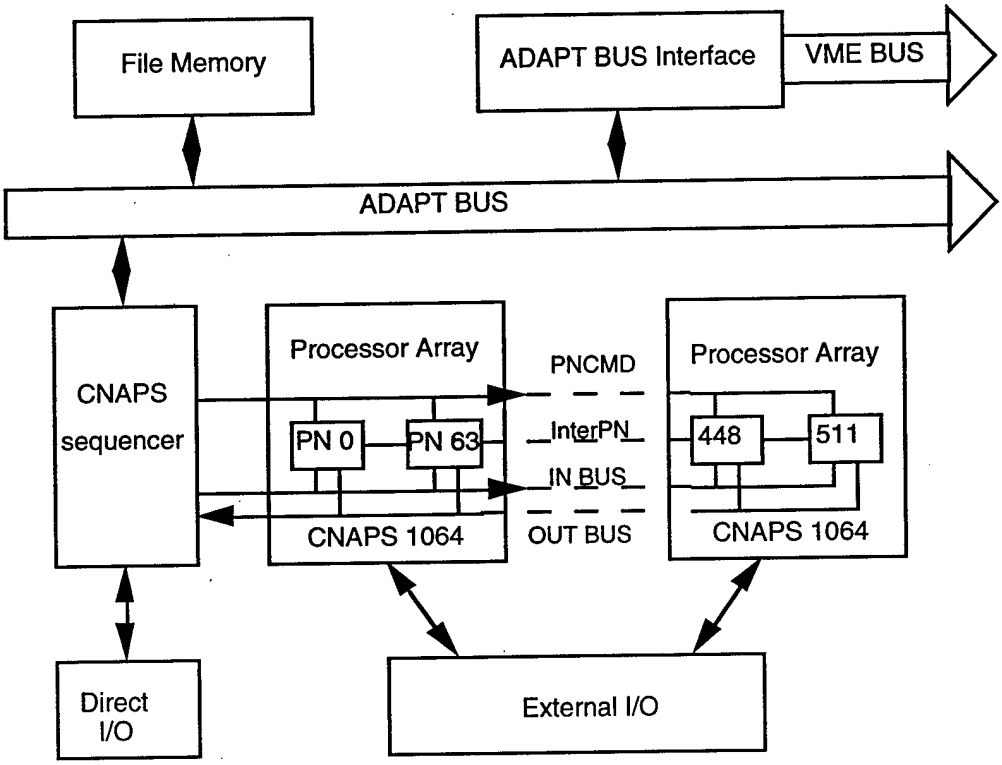

Figure 7.1: The CNAPS architecture for a 512 processing node (PN) array

We have implemented the CFAR algorithm on the CNAPS using the CNAPS-C language, a parallel-C environment. Though, neither optimized nor implemented at assembly language level,

this implementation still achieves a speed-up of roughly 10-26 times over a sequential-C implementation on the sparcstation for an image of 512X512 in size. The performance statistics are shown in Table 7.1. The speed-up we have quoted here includes I/O considerations. In the case of CNAPS, there is a feature known as Quick I/O which enables us to send the data directly to the processor chips if we can use the appropriate mezzanine boards.

Test image size	256 X 256
# of CNAPS processors used	256 (one row per processor)
Clock cycles used by the CNAPS impl.	CVCFAR: 690k cycles - 35msec OSCFAR: 1410k cycles - 71 ms
Clock cycles used by the serial impl. on Sparc 10	CVCFAR: 690k cycles - 861 msec OSCFAR: 1410k cycles - 774 msec

Table 7.1 - Performance Statistics of the CNAPS vs. Serial Implementation of the CFAR Algorithms

8. PHASE II OUTLINE

This Phase I has demonstrated the concept that wide-angle SAR can be effectively used to cue passive imagery for model-based recognition of targets. Our experience with the proof-of-concept and the insights we have gained during Phase I suggest that a fully developed Phase II prototype will advance the state of the art to reach the required level of progress in ATR technology for target acquisition applications. Even though, object recognition from FLIR imagery has been studied for well over a decade now, most of these efforts are less applicable to the air-ground target acquisition problem due to the large variation in target appearance from different view-angles. Meanwhile, very interesting results have been achieved by the model-based vision community. The notion of motion as an important cue is just beginning to gain momentum in the community even though we

proposed to use motion for ATR in 1989 and earlier [Raghavan et. al. 1991]. This Phase I work has reinforced some of these concepts.

The primary technical objective of the two-year Phase II work is to expand this concept of motion based ATR towards developing a prototype. The first year effort addresses several design issues including the multi-resolution based optical flow computing, perceptual grouping algorithms for hypothesization of model-image corresponding points, model-selection and indexing algorithms. Some of the tasks involved in the first year of the Phase II are outlined below.

- We would first undertake the completion of the multi-resolution based optical flow computation for eliminating the dependency of desired frame rate of the camera on the expected range of speed for targets in the image.

- We need to implement the perceptual grouping theory algorithms (step 2 shown in Figure 6) for establishing the correspondence between model-image points.

- The model-selection and indexing algorithms need to be (step 1 shown in Figure 6) implemented to enable automated selection of a model for computing the model-image transformation.

- A better error metric based on feature information (e.g., local orientation of edges) need to be developed to match the alignment error between the model and image points.

The focus of the second year effort is to develop a near-real-time prototype of the model-based ATR system by completing all the required peripheral modules designed in the first year. The following are some of the tasks involved in this stage.

- An end-to-end simulation of the prototype software system to emulate the model-based ATR system will be undertaken. This includes the placeholders for all the modules of the front-end motion information processor and the rear-end model-based vision component. Primarily, this simulation will take place on a serial machine like Sun Sparcstation.

- Parallelizable modules of the system (e.g., optical flow algorithm if we deal with video data) will be implemented on a parallel hardware platform. LNK currently has access to the CNAPS machine with 512 processors from Adaptive Solutions Inc. for real-time simulation of parallelizable algorithms. This machine has shown the potential for speed-up of the sequential simulations up to a factor of 100 or more in our current implementations related to image classification.

- A full scale demonstration of the final system will be accomplished with FLIR data acquired for ATR. This data will be exclusively used for the purpose of demonstration and will not be used for calibrating the system.

In addition to these major tasks required to complete the prototype, an important task of the Phase II work is to investigate the technology transfer. Some of the concerns in this regard include how the prototype system can be implemented on a stand-alone hardware which can be mounted aboard the aircraft, communication requirements, and other performance issues.

9. BIBLIOGRAPHY

Aguilera R.A., "Advanced IR Image Seeker Program", Proc. of SPIE, vol. 253, pp. 58-64, 1980.

Ballard D.H., "Generalizing Hough Transform to Detect Arbitrary Shapes", Pattern Recognition, vol. 13, pp. 111-122, 1981.

Barrilleaux, J.M., "A Biologically Motivated Algorithm for Image Interpretation Based on Multi-pass Multi-resolution Techniques," IJCNN, San Diego, 813-818, June 1990.

Battiti, R., Amaldi, E., and Koch C., "Computing Optical Flow Across Multiple Scales: An Adaptive Coarse-to-File Strategy", Int. Jl. of Computer Vision, vol. 6, no. 2, pp. 133-145, 1991.

Bhanu B., " Automatic Target Recognition: State of the Art Survey", IEEE Trans. on Aerospace Electron. Syst., vol. 22, no. 4, pp. 364-379, 1986.

Blake, A. and Zisserman, A., Visual Reconstruction, MIT Press, Cambridge, MA, 1987.

Brooks R. A., "Symbolic Reasoning Among 3-D Models and 2-D Images", Artificial Intelligence, vol. 17, pp. 285-348, 1981.

Brown, L.G., "A Survey of Image Registration Techniques", ACM Computing Surveys, vol. 24, no. 4, pp. 325-276, 1992.

Burton M. and Benning C., "Comparison of Imaging Infrared Detection Algorithms", Proc. of SPIE, vol. 302, pp. 26-32, 1981.

Chellappa R., Zelnio Ed., and Rignot E.,"Understanding Synthetic Aperture Radar Images", Proc. of Image Understanding Workshop, pp. 229-247, 1992.

Chellappa R. et. al., "On the Positioning of Multiple Sensors for Image Exploitation and Target Recognition", Proc. of IEEE, 1996.

Chin R.T. and Dyer C.R., "Model-Based Recognition in Robot Vision", ACM Computing Surveys, vol. 18, no. 1, pp. 67-108,1986.

Edelman S. and Weinshall D., "A Self-Organizing Multiple-view Representation of 3-D Objects", MIT AI Laboratory and Center for Biological Information Processing, Whitaker College, AI Memo no. 1146, August 1989.

Duda R.O. and Hart P.E., "Use of Hough Transforms to Detect Lines and Curves in Pictures", Communications of ACM, vol.15, pp. 11-15, 1972.

Fischler M.A. and Bolles R.C., "Image Understanding Research at SRI International", Proc. of DARPA Image Understanding Workshop, 1990.

Fogler R.J., Koch M.W., Moya M.M., Hostetler L.D. and Hush D.R., Feature Discovery via Neural Networks for Object Recognition in SAR Imagery, IJCNN, Baltimore, 1992..

Fonseca L. and Manjunath B.S., "Registration Techniques for Multi-Sensor Remotely Sensed Imagery", Photogrammetric Engineering and Remote Sensing, Sep. 1996.

Goto Y. and Stentz A., "The CMU System for Mobile Robot Navigation", Proc. of IEEE Int. Conf. Robotics and Automation, pp. 99-105, 1987.

Grimson W.E.L., "From Images to Surfaces" MIT Press, Cambridge, MA., 1981.

Grimson W. E. L. and Lozano-Perez T., "Model-Based Recognition and Localization from Spare Range or Tactile Data", Int. Jl. of Robot. Res., vol. 3, no. 3, pp. 3-35, 1984.

Herman M. and Kanade T., Incremental reconstruction of 3D scenes from mutiple, complex images, Artificial Intelligence, 30, pp. 289-341, 1986.

Heurtas A., Cole W., and Nevatia R., "Detecting Runways in Complex Airport Scenes", Computer Vision, Graphics and Image Processing, vol.51, no.2, pp. 107-145, 1990.

Heurtas A. and Nevatia R., "Detecting Buildings in Aerial Images", Computer Vision, Graphics, and Image Processing, vol. 41, pp 131-152, 1988.

Huttenlocher D. and Ullman S., "Recognizing Solid Objects by Alignment with an Image", Int. Jl. of Comp. Vision, vol. 5, no. 2, 1990.

Kamberova G. and Mintz. M., "Robust Multi-Sensor Fusion - A Decision Theoretic Approach", Proc. of DARPA IU Workshop, 1990.

Kanal L. and Raghavan S., "Hybrid Systems - A Key to Intelligent Pattern Recognition", Proc. of IJCNN, 1992.

Kienker P.K., Sejnowski T.J., Hinton G.E., and Schumaker L.E., "Separating Figure and Ground with a Parallel Network", Perception, vol. 15, pp 197-216, 1986.

Kohonen, T. Self Organization and Associative Memory, Springer-verlag, 1988.

Kosko B. Neural Networks and Fuzzy Systems, Prentice-Hall, 1990.

Lambird, B. et al, "Study of Digital Image Matching of Dissimilar Images," LNK Tech. Report ETL-0248, U.S. Army ETL, Fort Belvoir, VA, Jan. 1981.

Lambird B., Bailey G., and Lavine D., "Icue - Feature Extraction for Flight Simulator Databases", LNK Technical Report, Contract N61339-88-C-0049, Naval Training Systems Center, 1991.

Lavine D., Olson E., Lambird B., Berenstein C., Leifker D., and Kanal L., "Study of Digital Matching of Dissimilar Images", LNK Technical Report ETL-0385, U.S. Army, Engineering Topographic Laboratories, 1985.

Lavine D., Raghavan S., Gupta N., and Lambird B., "Neural Networks for Object Detection Using All-Source Imagery", LNK Tech. Report ETL-0585, Contract Number DACA76-90-C-0014, U.S. Army, 1991.

Leclerc, Y.G. Region Grouping using the Minimum-Description-Length Principle, Proc. DARPA Image Understanding Workshop, 1990.

Li H., Manjunath B.S., and Mitra S.K., "A Contour-Based Approach to Multi-Sensor Image Registration", IEEE Trans. on Image Processing, vol. 4, pp. 320-334, 1995.

Liow Y.T. and Pavlidis T., Use of shadows for extracting buildings in aerial images, CVGIP, 49, pp. 242 - 277, 1990.

Lowe D., "Three Dimensional Object Recognition from Single Two-Dimensional Images", Artificial Intelligence, vol. 31, pp. 355-395, 1987.

Lowe D., "Fitting Parameterized Three-Dimensional Models to Images", IEEE Trans. PAMI, vol. 13, no. 5, pp. 441-450, 1991.

Mallat S., A theory of multiresolution signal decomposition: the wavelet representation, IEEE Trans. PAMI, 11, pp 674-693, 1989.

Manjunath B. S. and Chellappa R., "A Unified Approach to Boundary Perception: Edges, Textures and Illusory Contours", USC-SIPI Report No. 167, 1991.

Marr, D., Vision, W.H. Freeman Press, New York, 1982.

Marr, D. and Poggio T., "Analysis of a Cooperative Stereo Algorithm," Biological Cybernetics, Vol. 28, 223-229, 1978.

McKendall R., "Statistical Decision Theory for Sensor Fusion", Proc. of DARPA IU Workshop, 1990.

McKendall R. and Mintz. M., "Non-Monotonic Decision Rules for Sensor Fusion", Proc. of DARPA IU Workshop, 1990.

McKeown D.M., The Role of Artificial Intelligence in the Integration of Remotely Senses Data with Geographic Information Systems. IEEE Transactions on Geoscience and Remote Sensing, Vol. GE-25, No. 3, May 1987, pp. 330-348.

Mohan R. and Nevatia R., Using Perceptual organization to extract 3-D structures, IEEE Trans. on PAMI-11, pp. 1121-1139, 1989.

Nagao M. and Matsuyama T., A Structural Analysis of Complex Aerial Photographs, Plenum, 1980.

Nevatia R., Price K. and Medioni G., "USC Image Understanding Research: 1989-1990", Proc. of DARPA Image Understanding Workshop, 1990.

Pomerleau D.A., Gowdy J. and Thorpe C.E., "Combining Artificial Neural Networks and Symbolic Processing for Autonomous Robot Guidance", Proc. of DARPA IU Workshop 1992.

Raghavan S., Bailey G., Lavine D., Lambird B., and Gupta N., "InFuse: Intelligent Sensor Fusion by Combining Neural Networks and Expert Systems", LNK Tech. Report for Contract N62269-90-C-0567, Naval Air Development Center, 1991.

Raghavan S. and Gupta N., "A Hybrid Approach to Extraction of Man-Made Objects from Multi-Source Imagery", LNK Final Report, Naval Surface Warfare Center, Contract N60921-92-C-0134, 1992.

Raghavan S. and Kanal L. , "Model-Based Vision via Model-Image Alignment", 2nd ATR Systems and Technology Conference, Ft. Belvoir, 1992.

Raghavan S., Gupta N., and Kanal L., "Computing Discontinuity-Preserved Optical Flow", Proc. of 11th IAPR Conference, Netherlands, 1992.

Raghavan S., Lavine D., and Kanal L., "Model-Based ATR Using Image Motion Cues from FLIR Image Sequences", SBIR Phase I Final Report, Contract No. F33615-93-C-1027, Wright Patterson Air Force Base, 1993.

Rajapakse, J. Acharya, R., "Multi-Sensor Data Fusion within Hierarchical Neural Networks," IJCNN, San Diego, 17-22, June 1990.

Roth M.W., "Neural Network Technology for Automatic Target Recognition", IEEE Trans. on Neural Networks, vol. 1, no. 1, pp. 28-43, 1990.

Rignot J.M. and Chellappa R., "Automated Multi-Sensor Registration: Requirements and Techniques", Photogrammetric Engineering and Remote Sensing, vol. 57, no. 8., pp. 1029-1038, 1991.

Seibert M. and Waxman A., "Adaptive 3-D Object Recognition from Multiple Views", IEEE Trans. on PAMI, vol. 14, no. 2, pp. 107-124, 1992.

Shufelt J. and McKeown D., "Fusion of Monocular Cues to Detect Man-made Structures in Aerial Imagery", Proc. of DARPA Image Understanding Workshop, 1990.

Smith G. and Austin J., Analysis of Aerial Images with ADAM, IJCNN, Baltimore, 1992.

Stewart C.V., "A New Robust Operator for Computer Vision: Theoretical Analysis", in Proc. of IEEE Int, Conf. on Computer Vision and Pattern Recognition, Seattle, WA., 1994.

Stockman G., Kopstein S., and Benett S., "Matching Images to Models for Registration and Object Detection via Clustering", IEEE Trans. on PAMI, vol. 4, no.3, 1982.

Thompson W.B., Pick H.L, Bennett B.H., Heinrichs M.R., Savitt S.L., and Smith K., "Map-Based Localization: The 'Drop-off' Problem", Proc. of DARPA Image Understanding Workshop, 1990.

Venkatesh S. and Rosin L., "Dynamic Threshold Determination by Local and Global Edge Evaluation", Graphics, Models, and Image Processing, vol. 75, no. 2, pp. 146-160, 1995.

Venkateswar V. and Chellappa R., A Hierarchical Approach to Detection of Buildings in Aerial Images, CS-TR-2720, University of Maryland, 1991.

Viennet E. and Fogelman F., Multiresolution Scene Segmentation using MLPs, IJCNN, Baltimore, 1992.

Vetterli M. and Herley C., Wavelets and Filter Banks: Theory and Design, IEEE Trans. on Signal Processing, V40, No.9, pp 2207-2232, 1992.

Weinshall D., "Model-Based Invariants for 3-D Vision", Int. Jl. of Computer Vision, vol. 10, no. 1, 1993.

Weiss I, "Geometric invariants and object recognition", Int. Jl. of Computer Vision, vol. 10, no. 3, 1993.

Wolberg G., Digital Image Warping, IEEE Computer Society Press, 1990.

Zhang X., Burlina P., Zheng Q., and Chellappa R., "Automatic Image-to-Site Registration", IEEE Trans. on Image Processing, 1996.

Zheng Q. and Chellappa R., "A Computational Vision Approach to Image Registration", vol. 2, no. 3,.pp. 311-326, July 1993.